TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

SONIA MATHEWS

# IMPROVING SPEND DATA QUALITY TO ENHANCE PROCUREMENT'S DECISION MAKING – A CASE STUDY

Master of Science thesis

Examiner: Professor Nina Helander and Professor Jussi Heikkilä
Examiner and the topic approved at the Faculty of Business and Built Environment Council meeting on the 8th of November

# ABSTRACT

**SONIA MATHEWS**: Improving spend data quality to enhance procurement's decision making – a case study
Master of Science thesis, 78 pages
November 2017
Master's Degree Program of Information and knowledge management
Major: Information, logistics and transportation
Examiners: Professor Nina Helander and Jussi Heikkilä

Keywords: spend data, spend analytics, procurement, P2P process, data quality, decision making, document analysis

Purchase-to-pay process is one of the organization's backbone as it enables purchasing different services and products to generate revenue. A lot of beneficial information is conducted in different steps of the process and one of the most important data set created is the spend data that reflects what has been purchased over time from which suppliers. This data consists of purchase order and invoice data. However, as the data amount has risen, there is a data quality issue. The challenge with vast data amounts is the poor quality that impacts direct to decisions that have been made upon poor data. Therefore, is critical to constantly assess the available data and see if it fits the purposes and targeted data quality level.

The aim of this thesis was to study how the spend data quality could be improved in the purchase-to-pay process for better decision making looking at data reliability, timeliness, completeness and accuracy. This thesis objective was to research what is the data quality in case organization's purchase order and invoice data and to compare them in price-quantity data. In addition, the thesis researched how different data was generated in various steps of purchase-to-pay process and how it affects decision making. The research methods used in this study was a case study and it had both qualitative and quantitative empirical parts and theoretical section. The quantitative research included analyzing purchase order and invoice data with document analysis method. There were 457 invoices analyzed and 301 purchase orders. In addition, qualitative data was collected by interviewing 23 different specialists, managers and leaders in the case organization in two separate interview rounds. The theoretical part of this thesis was based on scientific literature, though it was scarce on spend data and therefore this study aimed to build the gap between previous academic research and current literature.

Results of this study pointed out that there is a lot of room to improve invoice and purchase order data sets to achieve high quality spend data. This thesis results displayed that purchase orders and invoices commonly lack of quantity, item description and price. Furthermore, the study showed that invoice data sets are more complete and therefore should be utilized to have improved quality in price-quantity data. In addition, there were lot of challenges in the purchase order data, which were divided into three different categories: people related issues, process related issues and challenges that stem from too many alternatives. When spend data quality is good, people would also rely on the data and consequently, utilize the data more for various decision making situations and create a positive loop of data iteration. This would have a direct impact for improvements in many areas, such as creating savings, utilizing compliance vendors, improving due diligence and investing correctly.

# TIIVISTELMÄ

**SONIA MATHEWS**: Hankinnan kustannustiedon laadun parantaminen hankinnan päätöksenteon tueksi – case-tutkimus
Tampereen teknillinen yliopisto
Diplomityö, 78 sivua
Marraskuu 2017
Tietojohtamisen diplomi-insinöörin tutkinto-ohjelma
Pääaine: Liikenne, logistiikka ja informaatio
Tarkastajat: professori Nina Helander ja professori Jussi Heikkilä

Avainsanat: hankinnan kustannustieto, hankinnan kustannustiedon analytiikka, hankinta, hankintaprosessi, datan laatu, päätöksenteko, dokumenttianalyysi

Hankintaprosessi on yksi organisaation tärkeimmistä prosesseista, sillä se mahdollistaa erilaisten tuotteiden ja palveluiden ostamisen. Tämä prosessi tuottaa paljon tietoa prosessin eri vaiheissa, joista kriittinen on hankinnan kustannustieto. Hankinnan kustannustiedon tarkoitus on kuvastaa kuka on ostanut mitä miltä toimittajalta määrättynä aikana. Ostotilaukset ja laskut muodostavat hankinnan kustannustiedon selkärangan. Tiedon määrän kasvaessa alati, datan laatu on noussut ja näin ollen organisaatioilla on ongelmia pitää datan laatu tyydyttävällä tasolla. Heikko datan laatu vaikuttaa suoraan päätöksiin, jotka on tehty huonolla datalla. On tärkeää jatkuvasti arvioida saatavaa dataa ja tarkastella sopiiko se organisaation päämääriin ja tavoiteltuun datan laatuun.

Tämän diplomityön päämääränä oli tutkia kuinka hankinnan kustannustietoa voisi parantaa hankintaprosessissa, jotta saataisiin parempia päätöksiä. Tämän tutkimuksen tavoitteena oli tutkia mikä oli datan laatu case organisaation ostotilauksissa ja laskuissa ja vertailla niitä keskenään. Lisäksi diplomityö tutki kuinka eri dataa tuotettiin hankintaprosessin monissa vaiheissa ja miten tämä data vaikuttaa päätöksentekoon. Tässä työssä käytettiin tapaustutkimusta ja tässä oli kvantitatiivinen ja kvalitatiivinen empiirinen osuus sekä kirjallisuuskatsaus. Kvalitatiivinen osuus käsitti ostotilaus-ja laskutusdatan analysoinnin dokumenttianalyysillä. Yhteensä 457 laskua ja 301 ostotilausta käsiteltiin. Laadullista dataa kerättiin haastattelemalla 23 henkilöä, joiden roolit vaihtelivat asiantuntijasta, manageriin ai johtajaan. Työn teoreettinen osuus perustuu tieteelliseen kirjallisuuteen, jota oli melko vähän, jonka vuoksi tämä tutkimus pyrkii täyttämään tunnistetun tutkimusaukon.

Tämän työn tuloksena voidaan tunnistaa, että ostotilaus-ja laskudatalla on mahdollisuuksia parantaa datan laatua. Tässä tutkimuksessa havaittiin, että ostotilaus-ja laskudata ovat molemmat puutteellisia määrissä, kuvauksissa ja hinnoissa. Erityisesti, työ osoitti että laskudatassa on useammin määritettynä hinta ja määrä, joten tätä tietoa tulisi hyödyntää kun luodaan hankinnan kustannustietoa. Lisäksi työssä havaittiin haasteita ostotilausdatan suhteen, jotka jaettiin kolmeen eri kategoriaan: ihmisistä, prosessista ja vaihtoehdoista johtuviin haasteisiin. Kun datan laatu on hyvää, ihmiset luottavat dataan, ja tällöin käyttävät sitä – näin syntyy positiivinen datan iteraatiosykli. Hyvä hankinnan kustannustiedon laatu auttaa löytämään säästökohteita, parantaa eettisten sopimustoimittajien käyttämistä ja sopimushallintaa.

# PREFACE

It has been a long journey, to get me where I am now. One of my life's best times were spent in Tampere University of Technology. Especially my guild has provided unforgettable memories and taught me lot of lessons, outside of classes. University life has provided me skills to success in real life, not just the books and articles. I am really grateful of all the people I met there and the amazing spirit I was embraced with during those years.

I want to humbly show gratitude to the case organization, for the opportunity to conduct my master thesis. I want to thank all the interviewee's that have dedicated time and knowledge to support this research. I had a great, but challenging time working in the case company. Especially, I want to show gratitude for the closest colleagues that have lighten up my days with humor when I was frustrated. The case organization sincerely helped me towards my goals in many ways. I truly appreciate all the help and motivation Professor Nina Helander gave me and Professor Jussi Heikkilä for his strong guidance through the rough path of the thesis. It has been a pleasure to work with the both of you.

In addition, I want to address all my love for all the closest people to me. Many friends have supported and listened my ups and downs, as well as shared the agony of thesis process. However, I am most grateful for my family who have supported me through my whole life. You have taught me that anything is possible and trusted in me even in the stormiest weather.

Helsinki, 08.11.2017

Sonia Mathews

# TABLE OF CONTENT


1. INTRODUCTION ..... 1
   1.1 Research background and motivation ..... 1
   1.2 Research problem ..... 2
   1.3 Research scope ..... 3
   1.4 Research structure ..... 4
2. PURCHASING PROCESS ..... 6
   2.1 Purchasing model ..... 6
   2.2 Specification development ..... 7
   2.3 Vendor and market intelligence ..... 8
   2.4 Vendor selection ..... 8
   2.5 Drawing up the contract ..... 9
   2.6 Purchase ordering ..... 10
   2.7 Vendor evaluation ..... 10
3. DECISION MAKING AND DATA IN PROCUREMENT ..... 11
   3.1 The characteristics of making decisions in procurement ..... 11
   3.2 Decision making model ..... 12
   3.3 Spend analytics ..... 17
   3.4 Spend data ..... 19
   3.5 Data quality as a concept ..... 21
   3.6 Data quality dimensions ..... 22
4. RESEARCH METHODOLOGY AND EMPIRICAL STUDY ..... 27
   4.1 Research strategy ..... 27
   4.2 Research process ..... 29
   4.3 First round - preliminary study data, data gathering and analysis ..... 32
   4.4 Second round - study data, data gathering and analysis ..... 34
   4.5 Document analysis data, data gathering and analysis ..... 35
5. EMPIRICAL FINDINGS ..... 39
   5.1 The case organization’s purchase-to-pay process ..... 39
   5.2 The stakeholders in purchase-to-pay process ..... 41
   5.3 Current state of the data sets ..... 44
      5.3.1 Spend data ..... 44
      5.3.2 Purchase order data ..... 45
      5.3.3 Invoice data ..... 46
   5.4 Decision making in the case organization’s procurement ..... 48
   5.5 Summary of the impacts to spend data ..... 50
6. COMPARING PURCHASE ORDER DATA AND INVOICE DATA ..... 52
   6.1 Invoice data set ..... 52
   6.2 Purchase order data set ..... 53
   6.3 Comparison between purchase orders and invoices ..... 55
   6.4 Price-quantity data ..... 56

7. DISCUSSION AND CONCLUSION ..... 60
7.1 Main findings ..... 60
7.2 Theoretical and practical implications ..... 62
7.2.1 The P2P process ..... 62
7.2.2 The spend data quality ..... 64
7.3 Choosing purchase order or invoice data set for spend data ..... 66
7.4 Recommendations for the case organization ..... 67
7.5 Evaluation of the research ..... 71
7.5.1 Credibility and dependability of the study ..... 71
7.5.2 Confirmability and transferability of the study ..... 73
7.6 Suggestions for further research ..... 73
REFERENCES ..... 75

# LIST OF FIGURES AND TABLES

***Figure 1.*** *The scope of the thesis.*

***Figure 2.*** *Research structure*

***Figure 3.*** *The purchasing model (adapted from Day & Barksdale 1994; Fitzsimmons et al. 1998; Heikkilä et al. 2013, p. 31 van der Valk & Rozemeijer 2009; van Weele 2005).*

***Figure 4.*** *The elements of making a purchasing decision in procurement.*

***Figure 5.*** *Business intelligence process, adapted from (Choo 2002; Myllärniemi al. 2017; Vuori et al. 2008, p. 18).*

***Figure 6.*** *The relationships in identifying information needs and desires.*

***Figure 7.*** *Spend analysis' positive impacts.*

***Figure 8****. Components of spend data.*

***Figure 9.*** *Spend cube.*

***Figure 10.*** *The data quality dimensions adapted from (Ballou & Pazer 1985; Käpylä & Salonius 2013; Pipino et al. 2002; Wang & Strong 1996; Wand & Wang 1996).*

***Figure 11****. Chosen research methodologies (modified from Saunders et al. 2009).*

***Figure 12.*** *Research process.*

***Figure 13.*** *The document analysis structure.*

***Figure 14.*** *Case organization's P2P process.*

***Figure 15.*** *The different stakeholders in the P2P - process.*

***Figure 16.*** *Spend data attributes in the case organization currently.*

***Figure 17.*** *The current status of PO data in the case organization's ERP system*

***Figure 18.*** *Elements of invoice data in the case organization.*

***Figure 19.*** *Information needs, acquisitions and use in the case organization*

***Figure 20.*** *The roots of poor PO data.*

***Figure 21.*** *The relations between quantity (=Q), item description (=I D) and item price (=I P) in invoice data.*

***Figure 22.*** *The relations between quantity (=Q), item description (=I D) and cost center (=C C) in PO data.*

***Figure 23***. *Comparison between the theoretical purchasing model and the case organization's P2P process.*

***Figure 24.*** *Comparing the spend data elements in theory and empirical research.*

***Figure 25***. *Concept of spend data quality.*

***Figure 26.*** *Summary of the suggestions for the case organization.*

***Figure 27.*** *Evaluation of the research.*

***Table 1.*** *Invoice data attributes (adapted from Pandit & Marmanis 2008, p. 91).*

***Table 2.*** *PO data attributes (adapted from Pandit & Marmanis 2008, p. 5)*

***Table 3.*** *Summary of the data quality dimensions used in this thesis.*

***Table 4.*** *Preliminary interviews.*

***Table 5.*** *Empirical study interviews.*

***Table 6.*** *Attributes that are missing in invoices.*

***Table 7.*** *Attributes that are missing in the PO data set.*

***Table 8.*** *POs that miss quantity.*

***Table 9.*** *POs that miss quantity by category.*

***Table 10.*** *Invoices that miss quantity.*

***Table 11***. *Invoices that miss quantity by category.*

***Table 12.*** *The comparison between POs and invoice in quantity data.*

***Table 13.*** *Differences between purchase order and invoice data.*

## LIST OF ABBREVIATIONS AND KEY TERMS

Category = Where certain, similar commodities are mapped together

Commodity = A common name for either service or product

GL code = General ledge code

GR = Goods received

P2P process = Purchase-to-Pay process

PO = Purchase order

WBS = Work breakdown structure

Wild invoice = Invoice, that does not have a purchase order

# 1. INTRODUCTION

## 1.1 Research background and motivation

Procurement has been before a part of various functions such as supply chain management and sales. It has not been discussed a lot about, although it deserves its own spotlight. Pandit and Marmanis (2008, p. 5) explain that only in the recent years, has strategic procurement become a significant business process and van der Valk and Rozemeijer (2005) elaborate that even when making purchases, procurement is sometimes by-passed. The purchase-to-pay (later on P2P process) is a process that not much interest is given for due its supportive role and compulsiveness. Nevertheless, it is the backbone of finance and procurement as it enables doing business.

Within the service sector, procurement is especially a critical factor as purchasing consists of other elements rather than bulk buying for production. During past years, the significance of vendor relationships has been increasing at an explosive rate. Furthermore, selecting a dishonest vendor can make a huge negative impact on the company's wellbeing (Day & Barksdale 1994) and it can influence many areas such as sales, brand and customer loyalty. It is stated that company's profit can be increased by 10 % either by increasing sale volumes by 10 % or by decreasing purchase costs by 2 % (Lee et al. 2009). This been said, the spend of the company is directly linked to company's existence and success.

In many organizations, purchasing is seen as a more operational function than strategic or tactical (Van Weele 2005, p. 34), although many strategic decisions are based on spend data. Chowdary et al. (2011) argue that rarely companies have a proper process or resources to monitor purchasing processes, even though the procurement spend can rise up to billions of dollars in big companies. Spend analysis can cause positive impact in cost reduction, priorization of sourcing projects, compliance and monitoring the negotiated contracts and value (Pandit & Marmanis 2008). It gives insight which vendors are more valuable and what are the top commodities. Furthermore, with spend data, the planning and controlling of procurement improves as for an example sourcing managers are able to negotiate better contracts and financial department can estimate budgets more accurately.

The heart of spend management is the spend data (Pandit & Marmanis 2008, p. 6). Spend data consists often of merely invoice data, albeit it is not sufficient. To improve cost management, the spend data plays a vital role as it determinates the granularity, accuracy and coherency among many other important data quality dimensions. Even though it can be

argued that spend analysis is vital, there is a severe research gap as it has not been studied thoroughly after Pandit and Marmanis (2008) published a foundation for spend management, a book called "Spend Analysis – The Window into Strategic Sourcing". This thesis pursues to build a bridge to this century via an academic research and an empirical study in a form of a case study.

Correct data enables to build spend analytics – to really know where the money flow is going in the organization. Poor decisions are made if the used data is poor. The impact of poor data quality goes further than imagined, as it is estimated that costs are around 40-60% in a service organization (Redman 1998). It also affects intangible areas. Quality issues are rising as organizations seek for improved capability for analysis, not to mention the dilemma of larger and more robust data amounts (Hazen et al. 2014). The need for quality data grows exponentially as it eases people's working and improves efficiency. At the moment, an unbelievable 3,5 quintillion bytes of data is produced every day and the rate is growing all the time. Notwithstanding, it is claimed that only 0,5 % (Burn-Murdoch 2012) of the current data is been utilized and analyzed, making data visibility an issue in many organizations and the garbage in, garbage out has become a real challenge.

## 1.2 Research problem

The aim of this thesis is to research what is the current status regarding PO data and invoice data and what is the use for these data sets for the case company. It takes a stance on spend data quality that is at the moment build upon invoice data in the case organization. Furthermore, the study researches which data set would be beneficial to utilize POs as a data source for spend data. This thesis has a hypothesis that the PO data could be utilized for spend data. Therefore, it analyzes deeper the P2P process that creates the PO data. It also studies the spend data quality elements and how these elements are created. The thesis also discusses what kind of decisions are made in procurement characteristically.

The research problem of this thesis is:

**How procurement's decision making can be improved with enriched data?**

The research problem was partitioned into two questions:

- What kind of decision making is conducted in procurement and what information needs are there?
- What is the quality of purchase order data, invoice data and spend data in the case company?

These sub-questions were utilized to offer an understanding and framework for the research problem in a way that the different areas of examination can be identified and

answered. The first question surveys how decision making is conducted in procurement and what are the inputs. The second question dives into different data sets and quality as a concept. It offers a data quality study in the case company for PO, invoice and spend data.

## 1.3 Research scope

This study was made as an assignment for a large European listed service company. The primary focus area is procurement in the group level in the Nordic countries in IT relevant categories. Nonetheless, the research also studies strategic decision making in procurement– answering why is spend data in the heart of procurement. The foundation of the study is the P2P process as it frames the whole thesis (figure 1). Furthermore, the study investigates how decision making is conducted in the case organization and how the current data sets are utilized.

The case company had set a target to improve spend visibility. Particularly, the organization has pointed out that the problem roots from the lack of data availability, usability and poor data quality in procurement's data. This thesis focuses mainly on the latter problem. The most important issue is to solve the causing effect for poor data quality rather than implementing different temporary solutions as for an example data cleansing or many software systems built on top of enterprise resource planning (later ERP) system is not a stable solution in the long run.

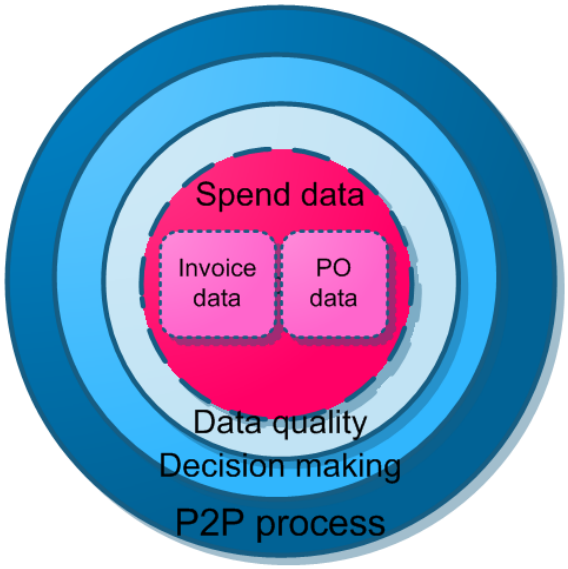



***Figure 1.*** *The scope of the thesis.*

The foundation of spend data is invoice data provided by vendors. Everyone in the company is allowed to make purchases varying from laptops to IT consultancy. The employees place the orders in the system, albeit the data quality challenge often stems from incomplete, wrong and incoherent data inputs from the users. This thesis takes in account

the various purchasing channels in the case organization. The case company manages the spend data with a category tree. Specifically, choosing the right category from multiple choices is extremely hard without a full training for purchasing. In this case, training all the employees thoroughly that may purchase rarely, is not a viable option, although these purchases also distort the spend data. Nevertheless, support for these purchases is provided but it is not sufficient. Regardless, this thesis does not take into consideration to change the extremely robust category tree, because it is an ongoing project already in the company. On the other hand, due to the data architecture, modifying ERP system is not an option due to its rigidness.

Therefore, this thesis compares which data set is more complete; the PO or invoice data to create improved spend data. The thesis researches what is the current spend data, invoice data and PO data in the case company as it is illustrated in figure 1. It conducts a comparison between invoice and PO data sets by mapping them manually.

## 1.4 Research structure

The research was divided into four major sections: introduction to the subject, the theory based on literature, research methodology and empirical research, empirical findings and the conclusion and summary part. This structure is presented in figure 2.

The second part presents a literature view based on previous research related to the subject. This section consists of two different chapters: the purchasing process and decision making in procurement. The purpose of the purchasing process chapter is to provide an understanding of the thesis context and procurement's main principles. The third chapter consists of what are the characteristics of decision making in procurement and how the decision making model can be applied to procurement. Furthermore, it provides insight on spend analytics, the opportunities it enables and the core of spend management, spend data. Last, quality dimensions in this field are depicted and what the usual errors are in spend data and how to tackle these issues with data quality.


***Figure 2.*** *Research structure.*

After the theory part, the thesis explores the research methodologies utilized in this thesis. It also considers in-depth explanation of the research process. The empirical part is systematic as it tackles the topic on two level; it has both qualitative and quantitative sections. First, it analyzes the case organization's status in qualitative way using interview data. The thesis evaluates the P2P process in the case organization and categorizes different decisions made in the procurement function. After this, the study utilizes document analysis to give insight to the PO and invoice data. In this part, a comparison between these data sets is made.

The thesis evaluates the spend data quality concept and whether to choose PO or invoice data set. It compares the different data sets. The thesis also takes a step back and looks at the possibilities the study brings to other organizations and academics. The last part is the summary of the thesis. Furthermore, it shows the main findings the study, theoretical and practical implications as well as suggestions for further research.

# 2. PURCHASING PROCESS

## 2.1 Purchasing model

Purchasing can be seen as a supportive activity in the value chain (Van Weele 2005, p. 12-15). The purchasing foundations is mostly discussed in van Weele's book Purchasing and Supply Chain Management (2005). It is one of the most vital elements in supply chain, often found in the upper stream. Nevertheless, support activity enables the primary functions (Porter 1985, p. 39). The reason for procurement being in support category stems from that it supports the primary activities (purchasing office supplies) and on the other hand, purchasing is a part of the end product as it consists of inputs which procurement has purchased. It must be remembered, that because of the supportive nature of procurement, the function has not been the key department to develop: companies prioritize enhancing first the functions that generate revenue.

In the literature, there are few different terms that are used for the same activity: purchasing, sourcing, procurement and supply chain (Van Weele 2005, p. 12-15). However, there are slightly different focuses in these definitions. Porter (1985, p. 39) describes procurement as purchasing consumable items and assets, that are either related to primary activities or support activities. In this thesis to purchase, to buy and to procure are used interchangeably. In addition, this thesis explores purchasing both physical products and services.

The purchase-to-pay process is also procure-to-pay process that is more common in literature, instead in this thesis the used concept is purchase-to-pay. Van Weele (2005, pp. 13; 46-53) proposes a general model that has six elements that is extended in van der Valk and Rozemeijer's (2009) service purchase model, adding two steps. The following purchasing model is utilized in this thesis (figure 3). The phases that are circled, are given the most importance in this thesis.

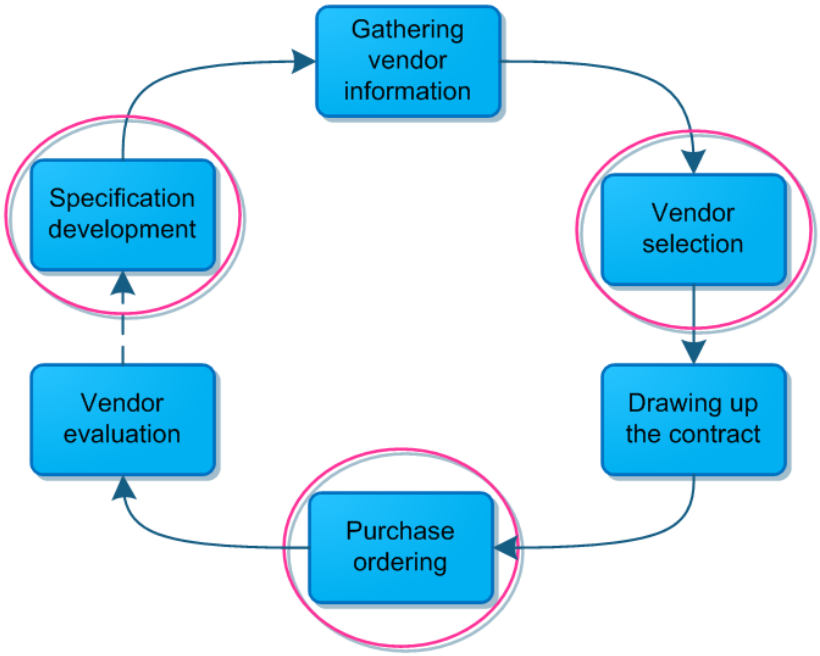



***Figure 3.*** *The purchasing model (adapted from Day & Barksdale 1994; Fitzsimmons et al. 1998; Heikkilä et al. 2013, p. 31 van der Valk & Rozemeijer 2009; van Weele 2005).*

On the other hand, Fitzsimmons et al. (1998) provide a purchasing process that is narrowed into four different phases. Day and Barksdale (1994) reflect the purchasing of services and decision-making process, leading to an eight step model. Heikkilä et al. (2013, pp. 30-44) determine a service purchasing model that consists of six different phases. The process can be subdivided into several strings of phases such as contract management, sourcing, strategic procurement, operational procurement (van Weele 2005, p. 13). It must be pointed out, that the original van Weele's (2005) process is not a continuous cycle, even though the following processes are (Day and Barksdale 1994; Heikkilä et al. 2013, p. 31; van der Valk & Rozemeijer 2009).

## 2.2 Specification development

In all of these models, the first step is to determine the specific requirements (Day & Barksdale 1994; Fitzsimmons et al. 1998; Heikkilä et al. 2013, p. 31 van der Valk & Rozemeijer 2009; van Weele 2005). In the first step, the do-or-buy (or make-or-buy regarding products) analysis is conducted and after that, the necessary specifications are drawn up (van Weele 2005, p. 47). Van Weele (2005, p. 47-48) illustrate the most important elements: quality specifications, logistics specifications, maintenance specifications, legal and environmental specifications and target budget. This initial stage is often underestimated and lacks thoroughness (Day & Barksdale 1994; Fitzsimmons et al. 1998; Heikkilä et al. 2013, p. 31; van der Valk & Rozemeijer 2009) and that is why Day and Barksdale (1994) uniquely propose two different steps to serve in more detail: problem recognition and defining purchase goals. Describing the problem answers questions like do we really need the commodity (Day & Barksdale 1994; Heikkilä et al. 2013, p. 31).

The purchase goals aim to improve the quality of client-provider relationship, not forgetting the final outcome (Day & Barksdale 1994). Fitzsimmons et al. (1998) have summarized problem definition, make-or-buy and scoping the specifications into a phase called need identification.

Similarly, van der Valk and Rozemeijer (2009) identify that specifying the service is one of the greatest challenges in purchasing services and they stress the value as their model has a detailed specification phase. Heikkilä et al. (2013, pp. 30-31) and van der Valk and Rozemeijer (2009) stress the fact that due to the broad variety of services, it is often very difficult to specify the services in great detail, compared to physical products. In van der Valk and Rozemeijer's (2009) model the detailed specification phase in particular highlights joining with the vendors to draw up the specifications together to achieve more value and detail.

## 2.3 Vendor and market intelligence

The second step is to gather information about the vendors (Day & Barksdale 1994; Fitzsimmons et al. 1998; Heikkilä et al. 2013, p. 31; van der Valk & Rozemeijer 2009; van Weele 2005, p. 49). Heikkilä et al. (2013) reference to supply market intelligence (SMI). Van Weele (2005, p. 51) also stresses the importance of knowledge of the industry.

Fitzsimmons et al. (1998) describe that information search encompasses personal contacts, references, recommendations and trade directory. Day and Barksdale (1994) emphasize the importance of this phase by dividing it into two steps: identifying the criteria list and after that, developing a short list of the possible vendor candidates. This thesis does not describe the criteria in detail as it is strongly dependable on the purchasing category. In the identification, references and recommendations are gathered if prior experience misses (Day & Barksdale 1994; Fitzsimmons et al. 1998). Furthermore, the prequalification list of the vendors is created (Day & Barksdale 1994). On the contrary, van Weele (2005, p. 49) has emerged the phase of gathering information and analyzing the vendors. Although, the elements are rather same as van Weele (2005, p. 49) lists phases such as preliminary qualification of vendors, listing vendors and sending requests for information (RFI). The request for information is actually its own step (van der Valk & Rozemeijer 2009). Many studies stress the vitality of collaboration and maintaining good relationships with the vendors (Day & Barksdale 1994; Fitzsimmons et al. 1998; van Weele 2005, p. 52).

## 2.4 Vendor selection

In the third phase, the vendors are analyzed and the one who meets best the vendor criteria list is chosen (Day & Barksdale 1994; Fitzsimmons et al. 1998; Heikkilä et al. 2013; van der Valk & Rozemeijer 2009; van Weele 2005, p. 49). Literature uses vendor evaluation and selection terms interchangeably, though there is a significant difference between

these phases. In this study, selection means how the vendor is selected and evaluation is a phase afterwards when the vendor performance is analyzed. This phase is rather same for physical products and services (van der Valk 2005). In this stage, the requests for information (RFI) (Heikkilä et al. 2013, p. 37) are sent based on the short list and after that request for quotation (RFQ).

Van der Valk and Rozemeijer (2009) prioritize this step also as a stepping stone for purchasing organizations. This stems from the fact that that measuring performance of service vendors is massively more complex (Day & Barksdale 1994; Fitzsimmons et al. 1998; Heikkilä et al. 2013; van der Valk & Rozemeijer 2009). The vendor assessment should be based on how well the vendors meet the expected criteria (Fitzsimmons et al. 1998) that is defined in the previous step. Equally important is to develop the process of selecting the vendor (van Weele 2005, p. 13), such as minimum and maximum requirements, prioritize the specification order and other rules. In addition, risk analysis is made.

Van der Valk and Rozemeijer (2009) clarify that when buying services soft criteria, such as trust and transparency, are more vital than hard criteria like price. Therefore, experience and references play a central role as numeric data in the request for quotation is not held in such great position. Additionally, comparing requests of quotations can be found difficult to assess (van der Valk & Rozemeijer 2009) as the specifications for the service in the first step often wave. It can be claimed, that the cornerstone of the whole strategic purchasing process is to define the necessary requirements in a sufficient level of detail.

## 2.5 Drawing up the contract

The fourth step concludes the vendor selection by drawing up the contract with the chosen vendor (Day & Barksdale 1994; Fitzsimmons et al. 1998; Heikkilä et al. 2013; van der Valk & Rozemeijer 2009; van Weele 2005, p. 49). Van der Valk and Rozemeijer (2009) take a high-level approach and depicts contract as a time period when chosen services or products are purchased and delivered. Many authors emphasize the significance of active contract management (Heikkilä et al. 2013, p. 37; van der Valk & Rozemeijer 2009; van Weele 2005, p. 56) and it leads to a more satisfied purchasing experience. Heikkilä et al. (2013, p. 39) specify that a good service level agreement (SLA) leads to decreased risk level. Creating the contract is regarding service business a lot more complicated and therefore also more essential. The SLA should include prices and terms of delivery, terms of payment, penalty clauses and other arrangements (van Weele 2005, p. 55).

First, the product or service should be defined thoroughly (Heikkilä et al. 2013; p.43). In addition, duration, goals, flexibility and performance indicators should be written in the contract. Van Weele (2005, p. 54-55) describes that drawing up the contract is industry-dependent, for instance considering legal matters. In addition, company cultures and policies have a high impact. A critical element of the contract is the pricing policy; whether

it is a fixed price, cost-plus or cost-reimbursable contract (van Weele 2005, p. 55). Another important issue is the payment terms. In practice, this is influenced by the industry competition and the vendor relationship; does the purchaser or vendor have more negotiation power. The contract should also include a statement of penalty clauses and warranty conditions. In addition, a clause of other arrangements such as insurance and terms of delivery should be written in the contract (van Weele 2005, p. 55).

## 2.6 Purchase ordering

As only van Weele (2005) touches this phase of the purchasing process, it can be said that this topic is limitedly researched in procurement's point of view. Purchase order data is generated in this phase. Van Weele (2005, p. 57) explains that in some situations, the contract acts as a purchase order. Purchase orders are initiated through purchase order requisition or materials requisition. For products, especially in inventories, van Weele states that a message could be generated from the system to remind procurement organization. Van Weele (2005, p. 57-58) discusses that vendors are countable to send order confirmations albeit the reality is often different. Last, vendor should sent an invoice matching to the PO. The actual invoice forms the invoice data.

## 2.7 Vendor evaluation

This step is twofold as it contains the evaluation of the vendor relationship as well as the assessment of the final outcome (Day & Barksdale 1994; van der Valk & Rozemeijer 2009). Mainly assessment is done based on the agreed criteria. Vendor evaluation is often based on the purchaser's or the procurement organization's expectations – if the vendor exceeded the expectations or failed to meet them (Day & Barksdale 1994). The evaluation should also include documenting the purchase and possible projects linked to it. With regard to the contract, excess work, warranty claims and penalty clauses should be discussed (van Weele 2005, p. 62).

The basis of evaluation differs depending on the commodity. Furthermore, assessing the quality of services is rather difficult, as there rarely are elements that would be easy to evaluate for the purchaser (Day & Barksdale 1994). For an example, how to measure how good quality the service is, is a lot harder question regarding services than physical products. For example, catering could be analyzed with dimensions such as taste of food, promptness and service friendliness. Considering physical products, the evaluation is more straightforward: was the product capabilities sufficient, was the quality good. Evaluation is often very subjective due to perceptions and so measuring is not adequate, but still commonly utilized. Overall, it is difficult to assess the added value and the relation to costs (van der Valk & Rozemeijer 2009).

# 3. DECISION MAKING AND DATA IN PROCUREMENT

## 3.1 The characteristics of making decisions in procurement

In literature, the description for decision making is quite wide and encompasses many different concepts and definitions. Similarly, it has been researched for quite long time now (Gilad & Gilad 1986). Therefore, exploring the particular angles of information and decisions regarding procurement is important. To elaborate, there are several decisions made in procurement, nonetheless literature stresses making the purchasing decision. The deciding elements that have an impact in purchasing process and should be considered are picturized in figure 4. As the money amount involved increases, often the strategic importance of the purchase increases concurrently. Risk rate escalates, too. For these reasons, also the responsibility of procurement grows as disciplines must be involved.

Making decisions to purchase is twofold (Day & Barksdale 1994; Fitzsimmons et al. 1998; van der Valk & Rozemeijer 2009). When buying services, there are situations that vary upon. Purchasing services stretches the challenge of having poor data that leads to bad decisions (Loshin 2013, p.3), as all the information needs might not be acknowledged by the purchaser. Nonetheless, the success of decision making in organizations relies highly upon the identification of needs that leads to the right information (Loshin 2013, p. 36; Pirttimäki 2007, p. 59). However, if there is lot of information it causes misunderstandings (Loshin 2013, p.3). For instance, if there is a vast amount of vendors that provide the same offering (Day & Barksdale 1994), the information load might increase drastically.

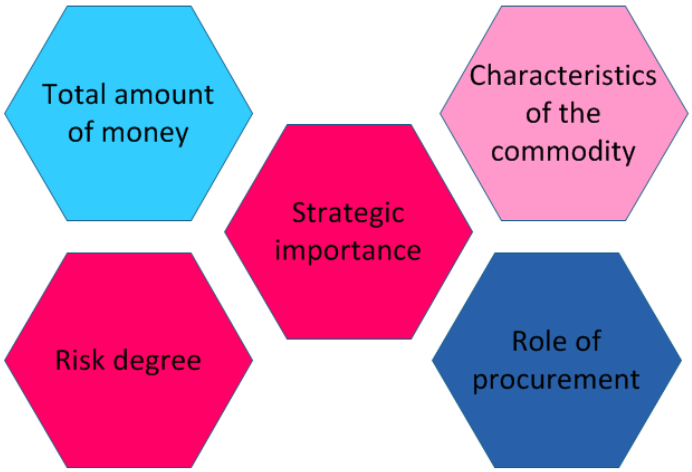



***Figure 4.*** *The elements of making a purchasing decision in procurement.*

Characteristic for procurement is that the decisions are often made by people who are not the ones who will work with the vendor (Day & Barksdale 1994). This means that the criteria to make the decision may differ from the operational needs. Top management might also see the purchasing decisions from another point of view, pointing out that services might be easier to buy than in reality (Day & Barksdale 1994; van Weele & Rozemeijer 2005). Pirttimäki and Hannula (2003) highlight that relevant information should be communicated effectively to decision-makers at all levels of an organization during a business intelligence process. After all, Pirttimäki and Hannula (2003) postulate that decision will become a more and more strategic in the future and organizations will understand that there are benefits to gain from empowering employees with applicable information.

Purchasing is highly affected by the formality and size of the organization. Van Weele (2005) clarifies that in an informal culture, following the purchasing process might have a low attention, depending on the strategic weight and risk level. As a result, the decisions are not made systematically or with the right information. Information should be useful and valuable knowledge or intelligence (Thierauf 2001; Pirttimäki 2007, p. 59) that is easily accessible. Vuori et al. (2008, p. 20) supports these views and highlights the refining of information. Vitt et al. (2002, p. 13) highlights the rate of speed in information – it should enable making decisions faster. To put it in other words, Eckerson (2007) defines operational decision making as utilizing information that it delivered quickly in the right operation. At the same time, procurement commonly has gatekeepers (van Weele 2005, p. 40) that prevent the transparent information flow from procurement to the vendor and vice versa. This disables the involvement of procurement and the possibility of organizational learning.

## 3.2 Decision making model

Decision making can be looked from a process-view and it can be placed into the purchasing model. The current literature presents different information process models or decision making models with only slightly different focuses (Choo 2002, p. 22; Gilad & Gilad 1986; Thomas J. 2002, p. 49; Pirttimäki & Hannula 2003; Vitt et al. 2002, p. 18). Day and Barksdale (1994) provide a classic model to make a purchasing decision, but the goal in this thesis is to explore also other perspectives of procurement decision making rather than merely purchasing. These variations mainly consider the process structure: cycle model and the number of phases. Furthermore, the collected data and the ways of storing the information differs, as well as the sources of information (Pirttimäki & Hannula 2003). There are few models that are theoretical such as (Choo 2002; Gilad & Gilad 1985; Myllärniemi 2015; Pirttimäki 2007), but the largest portion of these models are commercial, such as Microsoft's own decision making model (Vitt et al. 2002, p. 17-22) or the consulting company Thomas Group's developed model (2001, p. 48). The latter

ones are not as well applicable to procurement as this thesis discusses decision making in procurement in high-level.

The Microsoft's model consists of four different phases: analysis, insight, action and measurement (Vitt et al. 2002, p. 17-22) that comprise a continuous cycle. In general, information process models are quite often tailored for specific organizations to fit their need particularly (Pirttimäki & Hannula 2003). The Microsoft's model is a simplified version a decision making process. Nevertheless, Thomas Jr. (2001, p. 48) has developed a process model that includes six different phases, that are 1) planning and direction, 2) data collection, 3) information processing and storage, 4) analysis and production, 5) dissemination and last, 6) intelligence users and decision-makers. It is hard to define rigidly these models to different categories, as some are more business process performance frameworks or concepts than well-defined process models. This might be due to the fact that they are developed in businesses than stemming from academic research. As a result, the difference between these models is non-significant - in this way, can be overlooked.

Pirttimäki (2007) gathers together the most significant phases of decision making: defining needs, collecting information, processing information, sharing information and utilizing information. Myllärniemi (2015) depicts a process that closely follows the Choo's (2002, p. 24) model. It has five basic steps 1) defining information needs, 2) gathering information, 3) processing information, 4) sharing information and 5) utilizing information and feedback. Similarly, it is an ongoing-process. In an ideal process, the organization learns via feedback and adopts the information needs to fit the organization's business needs better. Myllärniemi (2015) takes also into consideration also the presentation and visualization of the data and the information systems behind.

One of the most well-known information process modelling is the Choo's (2002, p. 23-58) information management cycle. Most of the above mentioned models are fairly based on Choo's (2002, p. 24) model as it looks at the information process from a generic point of view. This model was chosen in this thesis as a cornerstone as it shows different sides of information processing; the abstract high-level but it can be also applied into operational or strategic level. Therefore, it suits well for procurement. Hence, some modifications were applied to provide a more comprehensive view. The continuous cycle consists of six phases (Choo 2002), figure 5:

1. Identifying information needs
2. Information acquisition
3. Organizing and storing information
4. Distributing information
5. Utilizing information
6. Adaptive behavior

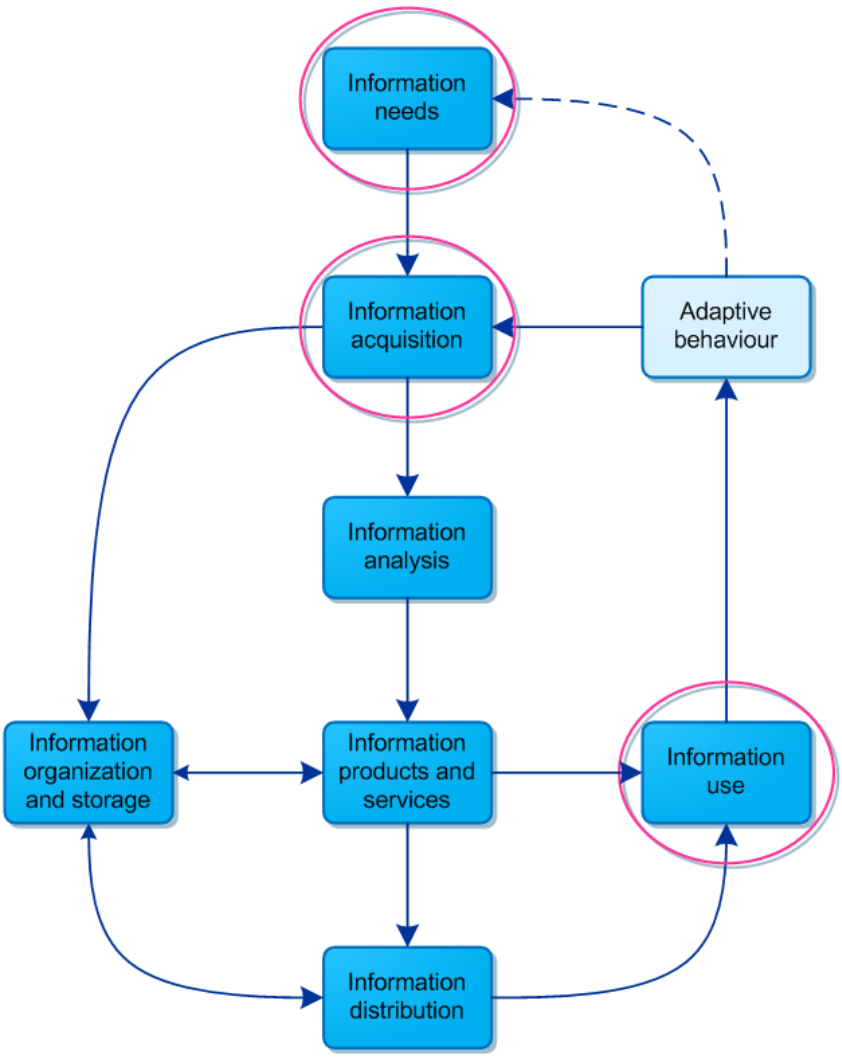



***Figure 5.*** *Business intelligence process, adapted from (Choo 2002; Myllärniemi al. 2017; Vuori et al. 2008, p. 18).*

According to Choo (2002, p. 24) it is possible to alter the phases and reduce the phases, if needed. Furthermore, Choo (2002) hypothesizes that all the information management process models do not include the identification of needs or information use. However, the models presented all had these elements (Thomas J. 2002, p. 49; Pirttimäki & Hannula 2003; Vitt et al. 2002, p. 18) so it can be debated that as the decision making process has evolved to a more standardized model. This thesis focuses on the first, second and last phase. Additionally, the stress of identifying information needs has increased in purchasing. Still, there lies a problem within the identification of business needs – solving a business need without a well-specified scope and stringent limits is a difficult task. Lining with these arguments, Pirttimäki (2007, p. 75) highlights the criticality of identifying information needs, making it the foundation of decision making. Vitt et al. (2002, p. 79) compliments this by reflecting that the information flow snaps without a thorough analysis of the information needs. The challenging part of this phase is the continuous change (Pirttimäki 2007, pp. 41-23, Vitt et al. 2002, p. 79). On the other hand, decision making

has many surprising elements that have to be solved so it can be claimed that there is a need for dynamic decision making.

Vitt et al. (2002, p. 15) present an analysis gap – the gap between what data is collected and what data is needed. Dervin (Choo 2002) presents a list of several information gaps, such as embeddedness: the options or the robustness of the options are unknown. Vuori et al. (2003, p. 31) exaggerate the difference between an information need and information desire. Acquiring information that is not meaningful, for example creating reports for the cause of routines, is a waste of resources. Vuori et al (2003, p. 31) compliments this by dividing information requests into three categories. These are 1. Information that is desired, but not adequate, 2. Information that is needed and desired and 3. Information that is needed but the need nor the desire is identified. The fourth category could be added: information that is needed but not requested. The relations of these elements and different information gaps are represented in figure 6. It is mandatory to prioritize different needs to maintain efficiency. With regard to procurement, this figure can be viewed as the circles represent different groups of people – the top management might request for information that the purchaser does not need. On the other hand, the operative end user might need information that he does not acknowledge beforehand.

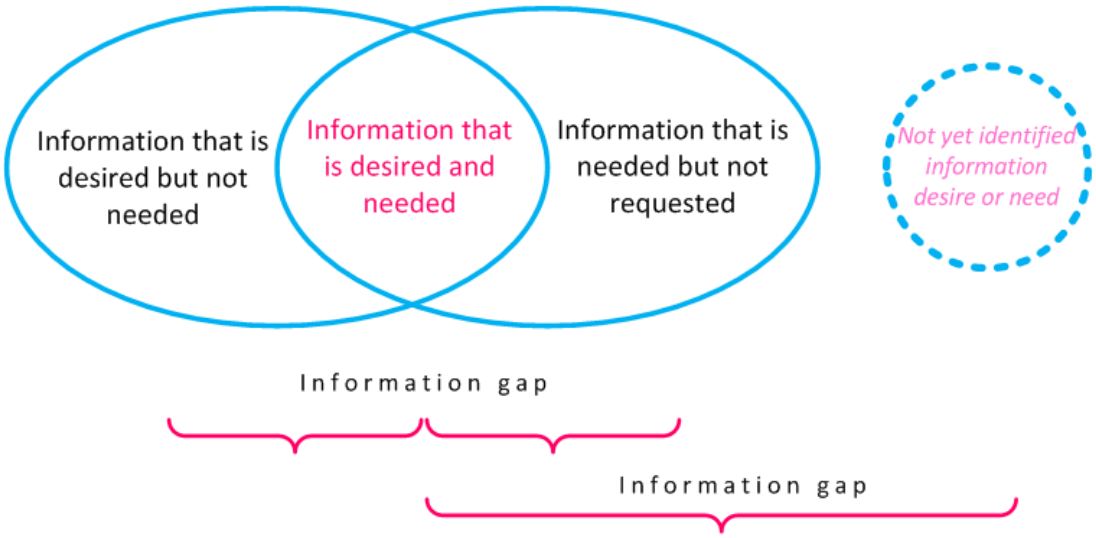



***Figure 6.*** *The relationships in identifying information needs and desires.*

Höglund and Persson (Pirttimäki 2007, p. 41) divide information needs into subjective and objective needs depending does it stem from an individual or the whole organization. Vuori et al. (2003, p. 33) claim that information need base on observations. Thus, this can be argued – many information needs are based on intuition, too, for example which vendor the purchaser feels comfortable communicating with. Vitt et al. (2013, p. 18) reinforce this statement and define a concept called mental model. This means the presumptions, observations and thoughts considering a certain subject.

Choo (2002) reflects the tight bond between information needs and information acquisition. The complexity of gathering information is discussed by many authors (Choo 2002; Pirttimäki 2007, p. 75; Vitt et al. 2002 p. 78). Pirttimäki (2007, p. 75) depicts that the gathered information can be quantitative or qualitative, collected from organization's internal data sources via primary or secondary research methods. These systems, such as ERP system, should be documented carefully to achieve a holistic information system architecture to ensure that the right data is collected. However, this thesis does not take a stance in technicalities, rather looks which data points can be collected where. In many organizations, information is scattered that makes the systematic information management challenging. An important source of information are the people themselves in the organization. This is often not regarded as high as it should, thus with humans there are emotion-related issues leading to resistance to change or the appeal for comfortability (Vitt et al., p. 25). It must be mentioned, that in practice is rarely impossible to acquisite all needed and desired information (Pirttilä 2000, p. 70).

After the analysis, different kind of information services and products are created as an outcome (Choo 2002). Information products vary from simple ones to extremely elaborated ones, from the short haul to the long run, to serve different organization functions to fulfil their targets. In procurement, common information products are contracts and spend data. These information products and services are stored and organized. Well-organized information products improve the organization's trust and operative efficiency (Choo 2002, pp. 25). At the same time, they are distributed to the organization as well as utilized (Choo 2002). They are the key items moving along the information process model. In an optimized situation, also tacit knowledge would be transformed into an information item – that would ensure that the information is not merely within the people. However, even the vendors do not commonly deliver all the documents they are supposed to (van Weele 2005, p. 37), which poses challenges. Nonetheless, managing all the data sets is a huge workload (Vitt et al. (2002, p. 79).

Sharing information happens through multiple channels, not matter which organization. Choo (2002, p. 42) stresses the goal of delivering the needed information to the right people precisely at the right place. In procurement, there has to be platforms for different stakeholders to share information, if it is not meant for the whole audience. Choo (2002, p. 42) supports this and states that finding the right channel that is appropriate for the target group is critical. When sharing information, the aspects of technical and social elements should be regarded (Choo 2002, p. 25). Thorough sharing demands both flexibility and promptness planning. Providing information enables the organization to learn and it is the prior goal of sharing information.

Utilizing the information is applying information in various decision making situations (Choo 2002). If this phase is not carried out well, the information lacks value – information only becomes valuable when it is utilized. It is critical that after information usage, the behavior changes. This should be measured, even though it is often claimed as default

as it determines the profit on the whole decision making process. In procurement, various other organizations also utilize the information to achieve the common goal to get more savings (van Weele 2005, p. 34).

The last phase concluded the change of the organization and its individuals' behavior as the understanding and knowledge increases. The achieved results affect the next iteration cycle (Choo 2002, p.18). This been said, the process starts from the beginning with slightly different information needs.

## 3.3 Spend analytics

Literature regarding spend data and spend analysis is very scarce. Thereupon this chapter is mostly based on Pandit and Marmani's (2008) book "Spend analysis – The Window into Strategic Sourcing", which describes the fundamentals of spend thoroughly. This been said, the existing literature has a research gap as it does not fulfill all current requirements, for example considering technology and IT as the book was written in 2008. The importance of spend has been debated, as Still et al. (2011) debate that spend management has been researched for years, whereas Pandit and Marmanis (2008, p. 3) discuss how spend has been overshadowed with budgeting. It might be that no constructed spend management was conducted, even though savings were on target. All in all, the studies of spend are rather limited and therefore, it can be concluded that the role of spend management was not high on priority list.

Chowdary et al. (2011) depict spend as the cost to the organization to generate revenue. The three main functions of spend analysis is to improve spend visibility, control and compliance (Pandit & Marmanis 2008). However, the causal connections between these elements can be debated. By sharing sufficient spend data, spend visibility is achieved. In addition, enhanced control over spend is a result of improved decision making that is achieved with spend analysis. This thesis focuses on spend visibility and impacts of spend analysis is illustrated in figure 7. Furthermore, spend analytics should include also visualization tools, reports, dashboards, feedbacks and scenario analysis (Pandit & Marmais 2008, p. 19). When there is a bunch of different transaction systems, lot of issues with the data, several vendors and thousands of transactions, the spend data quality is poor (Chowdary et al. 2011; Cox et al. 2005; Pandit and Marmanis 2008), so it requires a lot of resources before the value of spend analysis can be realized. However, via spend analysis there are lot of hidden gold-mines and Still et al. (2011) argue that the leading procurement organizations keep their position with advanced spend analytics.

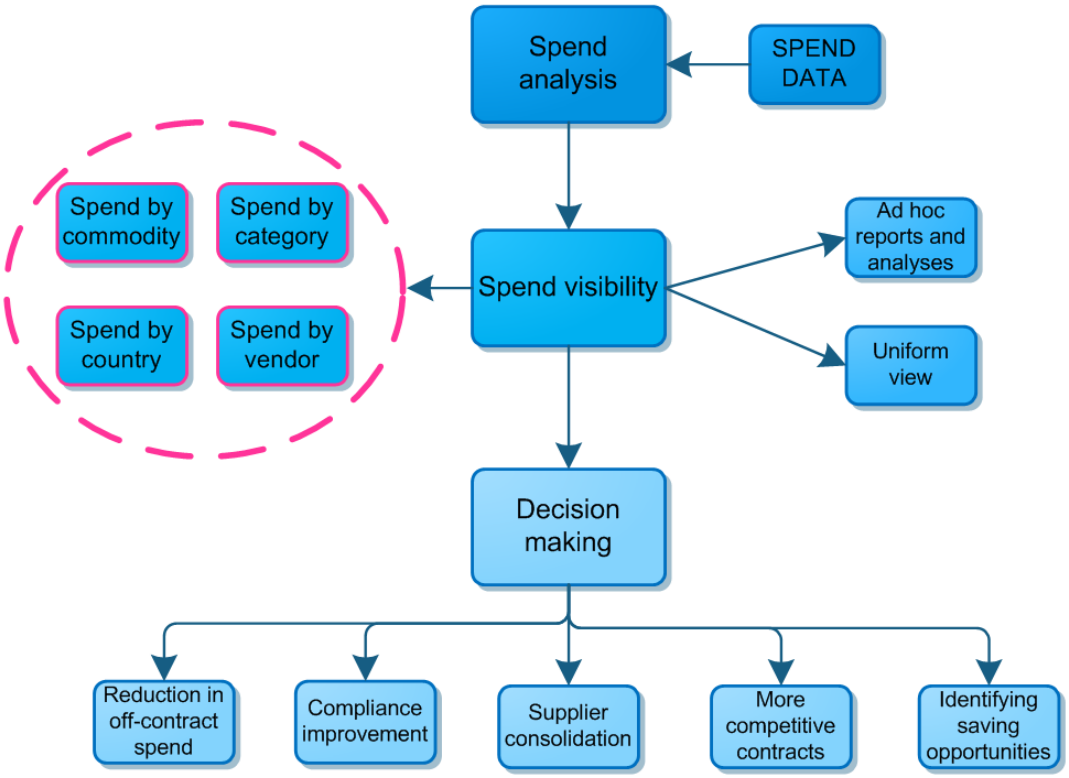



***Figure 7.*** *Spend analysis' positive impacts.*

By tracking and monitoring spend it is possible to recognize when there is spend leakage – lot of spend that is associated with nonpreferred vendors (Pandit & Marmanis 2008, p. 5). This usually stems from the fact, that with preferred vendors the negotiated terms and similarly, prices are lower. Furthermore, by aggregating these demands into one, there is a chance to achieve even better contract. Overall, with spend analytics the awareness of spending increases when there is information about spend by category, commodity, vendor or country. Often this information surprises the purchasing department (Pandit & Marmanis 2008, p. 5-6) in many ways, when shared and it leads to deeper analysis of the company-wide spend. There can be further discussions on topics like: "Why is this vendor used?", "Is this a strategic vendor, does it provide competitive advantage?", "What is the most purchased commodity, is it in line with the company strategy and could there be another vendor too?" This can provide possibilities to allocate spend in different way or insight to find saving opportunities (Chowdary et al. 2011; Pandit & Marmanis 2008). Spend analytics is a tool for the top management to control costs (Cox et al. 2002; Pandit and Marmanis 2008; Still et al. 2011) and it displays a spend trend if there is lot of sufficient spend data over the years. This is critical as it is impossible to manage if there is no knowledge of what the current situation.

Pandit and Marmanis (2008) and Chowdary et al. (2011) classify spend into three categories: indirect spend, direct spend and maintenance, repair and operations (MRO), whereas Cox et al. (2005) categorize MRO under indirect spend. In this thesis, the focus

is on indirect spend. Indirect spend consists of all supporting materials and services, such as business consultancy, IT hardware, insurance, marketing and office supplies (Pandit & Marmanis 2008). Indirect spend goes beyond the categorization of services and products (Mukherjee et al. 2008). The characteristics for indirect spend are maverick spend, distorted and huge vendor data base and suboptimal pricing (Pandit & Marmanis 2008). To clarify, maverick spend is the spend without a negotiated contract or agreements. As a result, there is no PO that leads to a so called wild invoice.

## 3.4 Spend data

Generally, spend data is historical purchasing data (Pandit & Marmanis 2008 p. 13), which is shattered in various systems. This data consists of PO data, invoice data and vendor data (Chowdary et al. 2011), whereas in this thesis the focus is on the first two data sets. Furthermore, Pandit and Marmanis (2008) state that spend data derives from many different business systems and applications. The data is often shadowed by finance and therefore not suitable as it is (Pandit & Marmanis 2008, p. 6).

Pandit and Marmanis (2008, p. 97) depict that there has to be two default attributes: the payment day and the paid amount. Without these two single data points, no spend data is possible to create. This information is usually in all transaction systems. Other critical attributes are service description, vendor name, cost center, commodity, division, geography and division (Pandit & Marmanis 2008, p. 97). These different dimensions are depicted in figure 8.

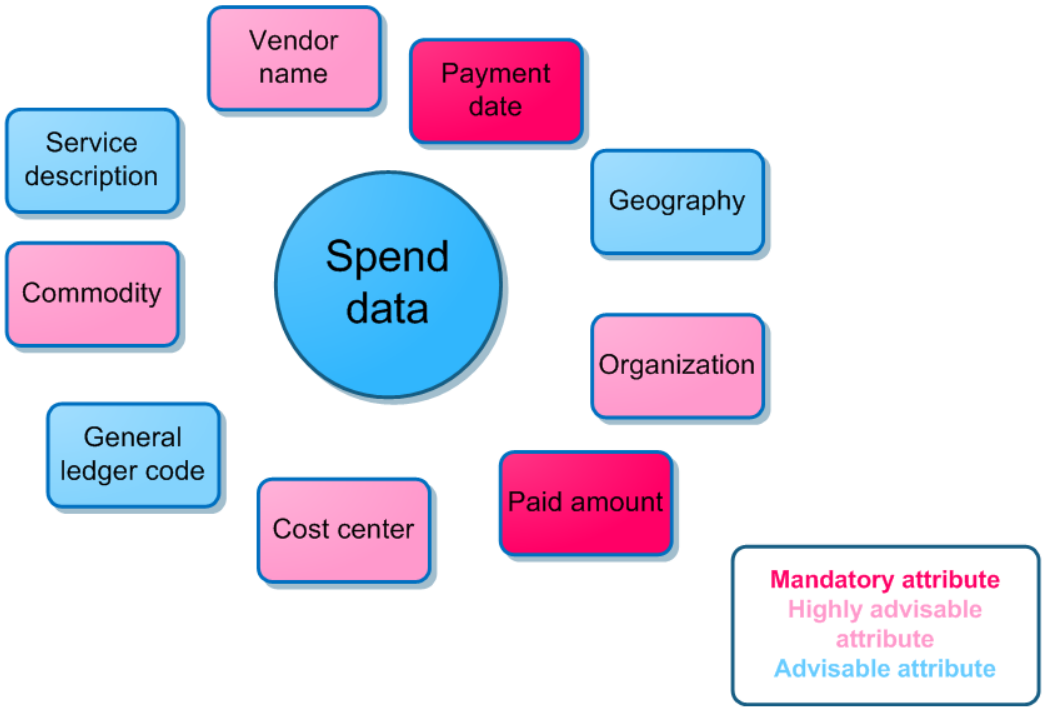



***Figure 8.*** *Components of spend data.*

The invoice data should contain various fields, most importantly vendor name, item price, invoice amount, invoice day and invoice amount. The whole list is displayed in table 1.

***Table 1.*** *Invoice data attributes (adapted from Pandit & Marmanis 2008, p. 91).*

| Invoice data attributes | |
|---|---|
| ***Vendor name*** | ***Taxes*** |
| ***Invoice amount*** | ***Surcharges*** |
| ***Invoice date*** | ***Discounts*** |
| ***Payment terms*** | ***Invoice number*** |
| ***PO number*** | ***Quantity*** |
| ***Due date*** | ***Item description*** |
| ***Item price*** | ***Total amount billed*** |

As the data is extracted from various systems, the data has different schemas and semantics (Pandit & Marmanis 2008, p. 15). The PO data has also several components that are listed in table 2. Chowdary et al. (2011) discuss that the raw data in PO data and invoice should obtain basic attributes like requestor, the amount of spend, PO number and date, product categories, vendor, geography, business units, vendor contracts and profit or cost center. General ledge code is mainly for accounting.

***Table 2.*** *PO data attributes (adapted from Pandit & Marmanis 2008, p. 5)*

| PO data attributes | |
|---|---|
| ***Vendor name*** | ***Cost center*** |
| ***Price*** | ***GL code*** |
| ***Creation date*** | ***Quantity*** |
| ***Payment terms*** | ***Item description*** |
| ***PO number*** | ***Total purchase amount*** |
| ***Approval date*** | |

Also van Weele (2005, p. 57) echoes these elements, but adds delivery time, delivery address and invoice address as PO attributes. It might be argued, that invoice address should be an element in the invoice data. However, this thesis merely surveys the attributes Pandit and Marmanis have researched and henceforth, does not take into account delivery matters.

Chowdary et al. (2011) state that there is a problem creating the drill-drown visibility because current analysis tools can only provide high-level dimensional data and cannot focus on many dimensions at the time. Pandit and Marmanis (2008) support this view by stating that the basic information, such as spend by category or vendor, is lacking in many

organizations. Chowdary et al. (2011) elaborate that in many organizations the spend visibility and control is high over only few core products or services. This been said, the real benefit of spend analysis is not utilized as the organization is not able point where exactly for an example return on investment level could be increased with a different investment strategy or better compliance, creating savings for the whole company. According to Pandit and Marmanis (2008, p. 16) there should be a possibility to drill-drown into vendors, commodities, cost centers, divisions, geographies and time. Figure 9 depicts the spend dimensions in a spend cube. With a cube, it is possible to examine the spend from various viewpoints at the same and the components can be changed to suit the purpose. On the positive side, it helps it with many issues. The dimensions that are missing can be located in ease, identifying the key vendors in the organization by spend and determining the main categories. It can also enable further analysis such as are there historical patterns in a selected category with chosen vendors.

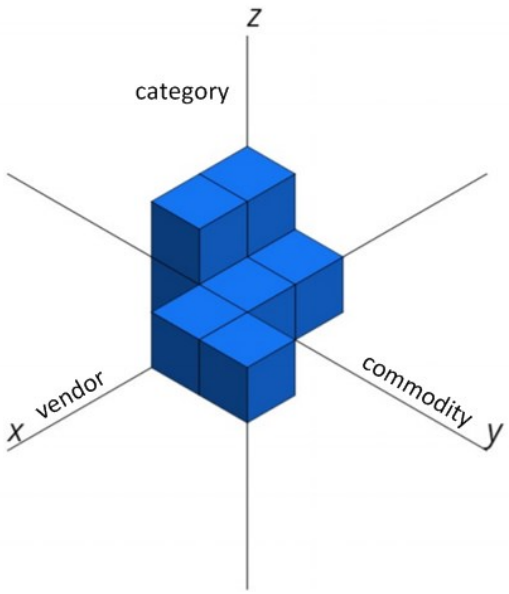



***Figure 9.*** *Spend cube.*

Still et al. (2011) debate that the spend analysis problem lies in the lack of consolidated spend database. Still elaborates the complexity of managing spend data as there is a bunch of data owners, multiple systems, unstable data architecture and a lot of manual data cleansing.

## 3.5 Data quality as a concept

Quality is a concept that has yielded numerous robust depictions and it has been researched widely for centuries (Ballou & Pazer 1985; Hazen et al. 2014; Redman 1998; Reeves & Bednar 1994, Wang & Strong 1996). Spend management has a high interest in this subject as it reflects strongly on the data set (Pandit & Marmanis, p. 13). Furthermore, data quality is the main driver affecting the question of data normalization and

enrichment (Pandit & Marmanis, 2008 p. 53), which is an issue in spend management. Additionally, the usefulness of the definition varies on several viewpoints: should quality be measurable (Reeves & Bednar 1998), what is the target group, and above all what is the purpose (Hazen et al. 2014). These elements define quality mostly.

Reeves and Bednar (1994) describe that quality has been depicted as value, meeting or exceeding customer's expectations and conformance on performance or specifications. Value is in the eyes of the beholder and in spend management it is the accuracy of spend data (Pandit & Marmanis 2008, p. 52-53). Nevertheless, Hoyle (2009, pp. 26-27) debate that if the customer's expectations are exceeded, it is considered as competitive advantage. Reeves & Bednar (1994) state that the model of quality as conformance on specification is the most precise depiction. It can be implied that the definition of quality is not affected by the time period (Redman 1998; Reeves & Bednar 1994; Wang & Strong 1996) as studies show that is has stayed the same for a long time and still no all-encompassing definition exists.

Literature stresses the definition of quality as fitness for use (Hazen et al. 2014; Tayi & Ballou 1998; Wang & Strong 1996) when looking at merely data. Data quality is also considered as a rigorously contextual model in which the chosen data quality might not be sufficient in another case (Tayi & Ballou 1998) and in this way, it does not fulfill the definition of fitness. For the above mentioned reasons, in this thesis data quality is illustrated according to Wang and Strong (1996) as data that is fit for use by the data consumers – in this case, the case company's employees in procurement. Notwithstanding, it is worth highlighting that every quality definition has its strengths and weaknesses (Redman 1998). To enumerate, fitness for use has little guidance for managers and the determination of use depends on the context. Equally, the data quality is greatly dependent on processes where the data is generated (Wand & Wang 1996). This been said, spend data is characterized by data coming from multiple IT systems and different processes with various purposes (van Weele 2006; Pandit & Marmanis 2008, p. 13).

## 3.6 Data quality dimensions

Several studies suggest that data quality is comprised of multiple dimensions (Ballou & Pazer 1985; Hazen et al. 2014; Käpylä & Salonius 2013; Pipino et al. 2002; Wang & Strong 1996; Wang & Wand). In this way, data quality can be divided into smaller sections and understood attribute by attribute. Ballou and Pazer (1985) were the first researches to introduce the four most recognizable dimensions of data quality: accuracy, timeliness, completeness and consistency. Many studies show that organization's focus too much on the accuracy attribute (Ballou & Pazer 1985; Wang & Strong 1985; Wand & Wang 1996), forgetting the other, equally important dimensions. Within these dimensions, there are deviations how these dimensions are defined (Wand & Wang 1996).

Nevertheless, the key to understand and create value through these dimensions are the rigorous depictions as the dimensions itself only vary a little and are easily confused. In general, there does not exist a well-defined set of data quality dimensions (Wand & Wang 1996), rather there is a fuzzy cluster of different terms. In Wang and Strong's (1985) research there is a categorization of the dimensions into four different groups: intrinsic, contextual, representational and accessibility. Wand and Wang (1996) focus more on the state-based model; what is the status in the information system; rather than data-based to provide a more objective view of the quality dimensions.

Generally, organizations wish for a single aggregate measure which upon it could be possible to evaluate the current data quality (Pipino et al. 2002). This kind of index would be beneficial (Pipino et al. 2002), thus Ballou and Pazer (1985) argue that it would be very subjective for the managers. Pandit and Marmanis (2008, p. 45) depict that there has to be two default attributes: the payment day and the paid amount. First, the payment day might have been formatted in various ways and second, currency affects the paid amounts. Therefore, these data elements represent the basic measurement of spend data. With a common measure, such as Dow Jones in the stock market, it would be possible to benchmark with other companies and to recognize weaknesses and strengths, as Ballou and Pazer (1985) claim that there is an obvious challenge to recognize the efforts that should be made to achieve improved data quality. The challenge within data quality is the fact that it has to be measurable, otherwise it cannot be managed. For instance, it is hard to measure free text dimensions such as item descriptions.

In this study, the subjective and objective dimensions are not separated. All the gathered data quality dimensions are illustrated in figure 10. The dimensions that cause most impact in spend data have been discussed in more detail below.


***Figure 10.*** *The data quality dimensions adapted from (Ballou & Pazer 1985; Käpylä & Salonius 2013; Pipino et al. 2002; Wang & Strong 1996; Wand & Wang 1996).*

**Reliability.** Agmont and Ahituv (1987) suggest that data reliability should be based on the data items relationships such as redundancy – are there duplicates of same data in different form or contradiction – are the items inconsistent. Pandit and Marmanis (2008, p. 106) support this view redundancy explaining that words can mean same even though they are written differently, for an example a status "complete" can be formatted as "100%" or "ready". Indeed, regularly item descriptions have poor quality that cannot be coded properly or there might be duplicates (Pandit & Marmanis 2008, p. 13). The key for reliable data is to validate the data (Agmont & Ahituv 1987). Thus, Agmont and Ahituv (1987) encompasses believability with reliability. According to Pandit and Marmanis (2008, p. 13) emphasize that only with good spend data quality there can be reliable analysis. Wand and Wang (1996) discuss that reliability indicates can the data be counted on and does it reflect the right data. Käpylä and Salonius (2013) rise questions regarding data sources, reputation and the purpose of the data. The latter one is one of the major hindrances regarding spend data, as there is no data source designated for spend data, it is rather a collection of different data sources. In Wang and Strong's (1996) research objectivity, believability, reputation and accuracy is under the same category.

**Accuracy.** Wand and Wang (1996) debate that inaccuracy reflect to other data deficiencies and they categorize accuracy with precision. Accuracy is associated with how precise the data is (Käpylä & Salonius, 2013; Wand & Wang 1996). Nevertheless, Pandit and Marmanis (2008, p. 111) argue that precision is how detailed the data is. For instance, the

data in the POs and contracts might not align (Pandit & Marmanis 2008, p. 91). Hence, Ballou and Pazer (1985) present a totally different point of view as they depict value as are the data points corresponding to the actual data. This conflicts with Pipino et al. (2002), because in their study the mentioned definition is determined as correctness. According to Pandit and Marmanis (2008, p. 106) misspelling also is an issue in spend data that leads to deviations in the data. Accuracy can be claimed to be entirely self-dependent (Hazen et al 2014). Pandit and Marmanis (2008) reflect to accuracy as the same term as data quality, which is a misleading perception. Accuracy often takes into consideration also reputable sources (Wang & Strong 1996), though Käpylä and Salonius (2013) categorize reputation to be an entirely own dimension.

**Timeliness.** This dimension is way more coherent with different studies timeliness represents how sufficiently up-to-date the data is (Käpylä & Salonius 2013; Pipino et al. 2002; Wand & Wang 1996). This might stem from the fact that timeliness is more general definition. In spend data, timeliness might be a problem as the spend data resides in multiple IT systems and there might not be proper integration (Pandit & Marmanis 2008, p. 13). The first definition was created by Ballou and Pazer (1985) that as "the recorded value should not be out-of-date". Käpylä and Salonius (2013) drill down to the fact, is the data accessible at the right time. Additionally, Wand and Wang (1996) state that untimeliness is a delay in real-time and when the data is accessible. Timeliness can be decomposed into two dimensions: currency and volatile (Hazen et al. 2014; Pipino et al. 2002; Wand & Wang 1996). Specifically, the rate of change - volatile - affects timeliness (Pipino et al. 2012; Wand & Wang 1996), because it is way harder to keep up with a rapid acceleration. Currency can be described as the length of time since the data was updated (Hazen et al. 2014). For timeliness, it is essential that the data is updated for the purpose (Käpylä & Salonius 2013; Pipino et al. 2002). The updating is difficult due to geographical limitations and division cultures and there might be laggardness from system to system in spend data (Pandit & Marmanis 2008, p. 13).

**Completeness.** In general, Ballou and Pazer (1987) break in the basic definition: the data has to be comprised with all the necessary values. In spend data, the data often lacks attributes and it makes it impossible to do reliable spend analysis with poor data. This can be described as the minimum ratio (Hazen et al. 2014Thus, Pandit and Marmanis (2008, p. 93) suggest uniting commodities with universal codes. Additionally, spend data might include several different currencies that has to be normalized (Pandit & Marmanis 2008, p. 93). On the other hand, Wand and Wang (1996) propose that data should display the necessary and meaningful elements in the information system. Similarly, Käpylä and Salonius (2013) stress how comprehensive the data is and how in depth has the subject been covered in the data. With regard to spend data, case organizations are often striving for enriched spend data because currently the data is concise (Pandit & Marmanis 2008, p. 13). The vendor lists would provide deeper knowledge if they had credit and quality ratings within. Moreover, even a higher viewpoint is provided as Wand and Wang (1996)

do not only relate to data concepts. Completeness reflects to the degree which entities and attributes are not missing from the data (Pipino et al. 2002). Pipino et al. (2002) divide completeness into schema completeness, column completeness and population completeness. In an Excel sheet column completeness can be depicted as missing values in the column and population that are there all the necessary rows.

**Relevancy.** Pipino et al. (2002) clarifies relevancy to describe how helpful and applicable the data. Likewise, Käpylä and Salonius (2013) mention the usability of information. It can be debated, is relevancy more of a result of good data rather than its causal connection with data quality. Notwithstanding, also Käpylä and Salonius (2013) highlights the importance of information in relevancy dimension. Similarly, Pipino et al. (2002) puts together concise representation and ease of manipulation to evaluate is the information outcome desirable or not.

**Consistency**. Consistency is an issue that lies in the core of spend data (Pandit & Marmanis, p. 13). To be clear, there is no particular definition for consistency as it is more a subjective dimension (Hazen et al. 2014; Pipino et al. 2002). Consistency can be defined as the redundant values across tables (Pipino et al. 2002) or the degree which related data match in terms of structure and format (Hazen et al. 2014). For an example, the spend data might be in several different languages (Pandit & Marmanis 2008, p. 13). Furthermore, Ballou and Pazer (1985) illustrate the dimension as the "representation of the data value is the same in all cases". Pandit and Marmanis (2008, p. 13) verify that also vendor names deviate even though the vendor is the same. For instance, a Company X can be named as Company AB, Company, Company Oy or by a business division. Additionally units might differ (Pandit & Marmanis 2008, p. 106). Hazen et al. (2014) decompose consistency into intra-related (a range of possible values) and inter-related (presentation within the same structure). This can be easily mixed up with consistent presentation. The dimensions are summarized in the table 3 below. These definitions are utilized in this thesis.

***Table 3.*** *Summary of the data quality dimensions used in this thesis.*

| Data quality dimension | Definition |
|---|---|
| ***Accuracy*** | Data is precise and correct. |
| ***Reliability*** | Data can be relied upon and reflects the desired data. |
| ***Timeliness*** | Up-to-date data is accessible at the right time. |
| ***Completeness*** | All the necessary attributes and entities are in the data. |
| ***Relevancy*** | The applicability and usability of the data. |

# 4. RESEARCH METHODOLOGY AND EMPIRICAL STUDY

## 4.1 Research strategy

To understand the philosophical assumptions and positions is critical as they have a great impact on how the research is conducted. Based on Saunders et al. (2009, p. 9) this thesis is an applied research, that aims to produce a solution to problem. It has objectives from people outside and a tight time scale. (Saunders et al. 2009, p. 9). In this thesis, the approach is adopted from Saunders et al. (2009, p. 109) onion model, which is presented in figure 11. The model is built like blocks – the foundation of a research is the research philosophy, approach and strategy. After these decisions, the choice of methods and time horizon is made. The uppermost layer illustrates the techniques that are utilized in the research. (Saunders et al. 2009, pp. 109-110)

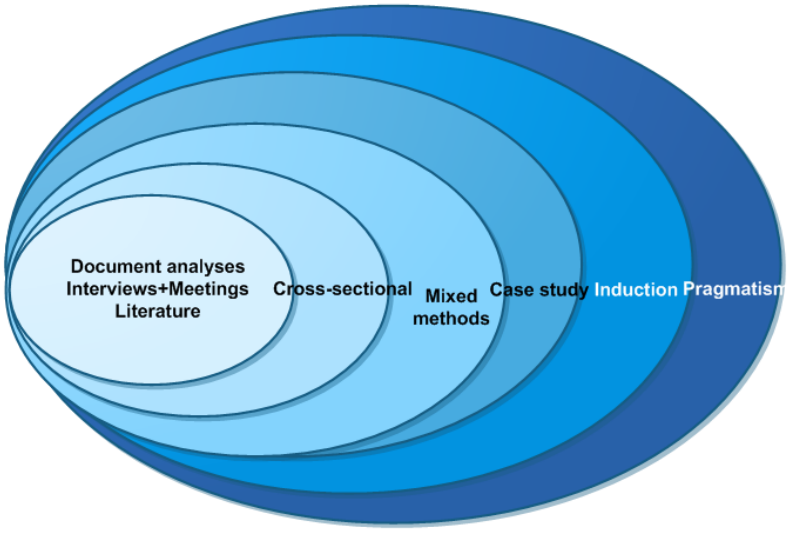



***Figure 11.*** *Chosen research methodologies (modified from Saunders et al. 2009).*

The research philosophy decision should be based on the restrictions about validity and realibity about the different philosophies (Saunders et al. 2009, p. 109), not forgetting the ontology and the basic assumptions. Every researcher has his own values, perception of life and associations that reflect strongly on the study, albeit if an objective or neutral approach is desired. Pragmatism is a research philosophy that is adaptable for different parts of the study and looks at information from the practical viewpoint (Saunders et al. 2009, p. 109). This means that the empirical and theoretical material can be regarded as a tight spider web to answer the ambiguous research question. In addition, pragmatism enables observations and subjectivity (Saunders et al. 2009). Because the research prob-

lem is also an organizational issue, it is important to understand the existing organizational culture and behavior and on the other hand, it has a very strong analytical grip due to the hands-on empirical material. Moreover, in pragmatism mixed methods are common and in this thesis, both quantitative and qualitative data is used for pursuing a holistic view of the problem.

Theory of information process models and data quality and its frameworks is a rather lot researched topic. Conversely, placing this theory in procurement's context is unusual and has been very narrowly studied in previous researches. Especially spend data is the core element in this thesis and it has not been researched almost at all. It can be stated, that there is a research gap. Therefore, this thesis has an inductive approach as it pursues to build new theory (Saunders et al. 2009, p. 125). The uttermost important fact is that the research question and topics elaborate within the research process and the study has an iterative, almost agile research approach. Saunders et al. (2009, p. 127) also emphasize gaining a close understanding of the context. It must be mentioned, that induction consumes more time and the final outcome is more unstable (Saunders et al. 2009, p. 127). Time is a restricted resource in this thesis, nevertheless other mandatory equipment were offered by the case company.

This thesis researches only one organization and its specified problem so the choice of a single case study was a natural one. Yin (2012, p. 10) states that if the research question is "how", case study is a suitable option due of its exploratory character. Saunders et al. (2009) confirm that the strategy is especially used when aiming for a rich understanding of a certain context. Yin (2012, p. 15) elaborates that case study is utilized when fieldwork is done and when observation plays a great role. In this thesis, participation in the case organization has a major impact. Yin (2012, p. 40) continues that case study strives to generalize the results to achieve high-level concept that can be applied in various situations. Also this study has an endeavor to be utilized in procurement, sourcing and purchasing not depending of the organization. Besides that, the research takes part in the discussion of data, data analytics and technologies. In an embedded case study, the focus is also in other departments besides the target department (Saunders et al. 2009), in this case the purchasing department. For this thesis, it is beneficial for an example interview people also outside the main function for a deeper understanding of the phenomena as it considers other units too. Similarly, the role of purchasing as supportive activity affects the choice of embedded single case study.

When using both qualitative and quantitative methods, there is a higher probability that there might be unpredictable solutions (Saunders et al. 2009, p. 153-154). Yin (2012, p. 88) illustrates the difficulties with collecting data: the ambiguous interviews and the special arrangements that might have to take in place. In this thesis, mainly primary data is used as the document analysis results in insights of the data that has been collected. In this case, no data of the researched topic was in the case company. This results that the purpose of the data collection can be altered particularly to this study (Saunders et al.

2009, p. 271). In addition, interviews support the thesis with primary data. Some case company documents are analyzed that have a limited availability only for employees in the organization. Like Saunders et al. (2009, p. 269) illustrate it is beneficial to compare data with the collected primary data.

Choosing mixed methods is convenient, but at the same it has own challenges due to the fact that different interpretations and methods can lead to a contradicted perception (Saunders et al. 2009). In this thesis, complementary design is utilized. When using a complementary design, there is a possibility to mix research strategies for an example interview to understand the reasons that where found with quantitative data (Saunders et al. 2009). In addition, in facilitation multiple data collection methods can be utilized (Saunders et al. 2009). Yin (2012, p. 119; 203) argue that a principle of case study is utilizing multiple sources and it provides alternative perspectives. The choice of cross-sectional research strategy was a natural choice, because in cross-sectional time horizon the study is conducted in a particular time in certain organization (Saunders et al. 2009). This thesis explores the current situation for the data sets and decision making in the case organization and pursues to provide alternatives for the future.

Document analysis is a method that utilizes different kind of documents in a systematic way (Anttila 1998). The documents are divided into secondary and primary documents. Anttila (1998) and Ojasalo et al. (2014, p. 136) explains that the spectrum of documents in document analysis is wide: web pages, diaries, reports, speeches and other written materials. In this thesis, mainly primary documents such as invoices and POs are researched. Anttila (1998) characterizes document analysis that is beyond straightforward and direct observations. The purpose of document analysis is to create a clear and informative description of the researched problem. It has four main phases: collecting the data, analyzing it, identifying the repetitive patterns and interpreting these findings (Ojasalo et al. 2014, p. 138). This study surveys the recurring dimensions of each data set in the data pool. Based on the document analysis, specific repetitive data attributes are analyzed, categorization is conducted as well as a comparison based on these data sets. Additionally, overall picture of the phenomena is enriched with details and a bridge between the theory and empirical part is built. The disadvantage of document analysis is that the documents have been previously created for other purposes and therefore are not straightforward to utilize. However, Ojasalo et al. (2014, p. 136) reflects that in positive light; the documents are then in their natural shape.

## 4.2 Research process

Saunder et al. (2009, pp. 10-11) depicts that commonly a research process includes formulating topic, literature review, research design, collecting and analyzing data. This thesis process is shown below (figure 12). The process an entity that is divided into two sections – the original topic and then the final topic. In the picture, all gathering tasks are

colored in pink, analysis in blue, literature review in grey and in light blue the last wave before the final stage. Everything regarding research questions is in red.

Figure 12 illustrates gathering knowledge is the first step of the elaborated process. The literature review and preliminary study data collection are done side by side. This is the foundation of the research. The preliminary study data analysis ties up the first and the second section.

In the second section, first the research questions are formed based on insights of section one. Studying literature and conducting practical preliminary study analysis are executed side by side to enrich the researcher's point of view. The literature is written iteratively through many other processes. Alongside with this also the second round of interviews are conducted to deepen knowledge of particular issues that stood out from the preliminary study analysis. The study data is then analyzed with the preliminary study, so that they are tied together and create a whole picture. The collection of document analysis data was a long process that took place during the process of secondary data.

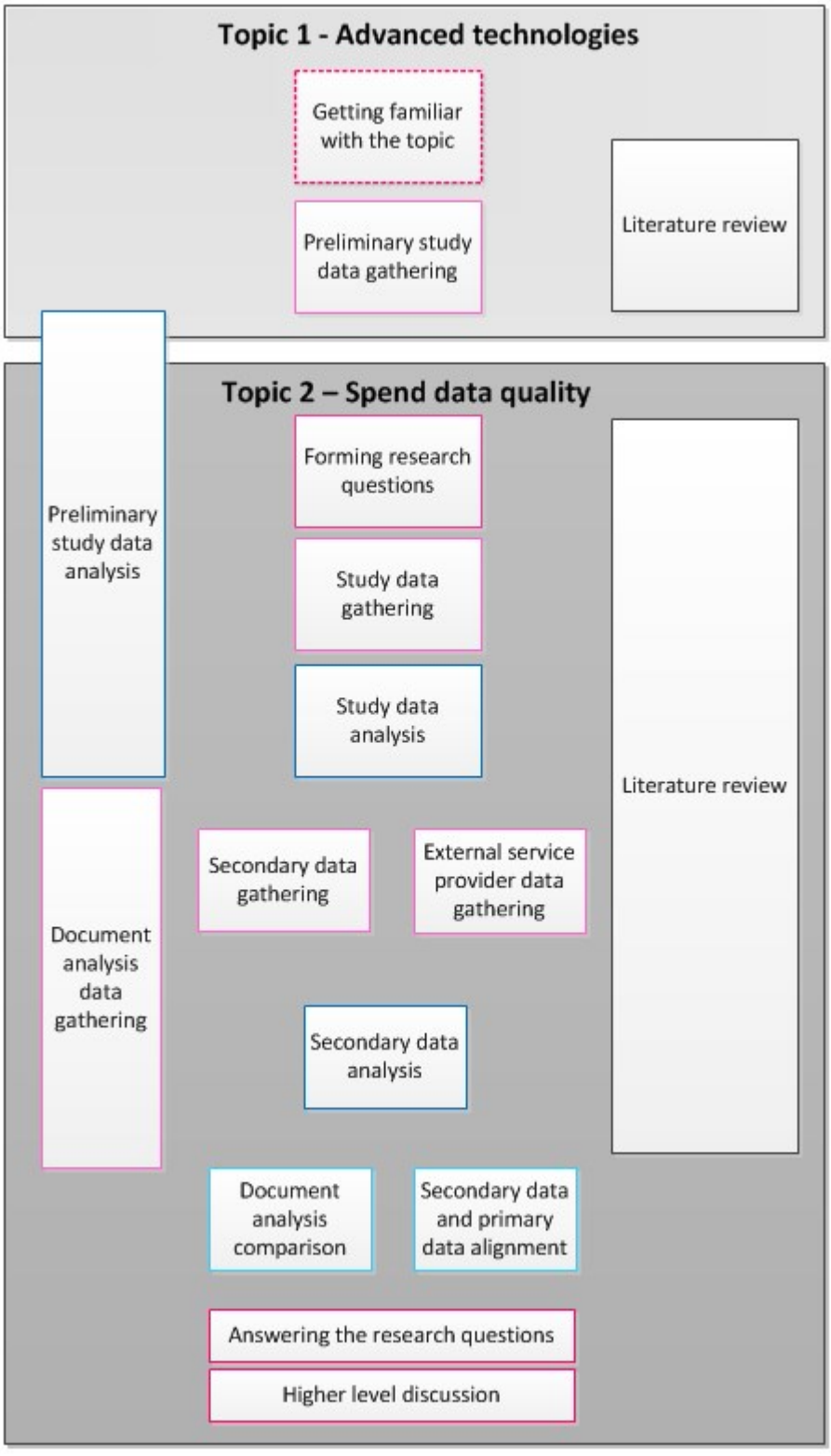



***Figure 12.*** *Research process.*

In the last phases, the invoice and PO data are compared in document analysis comparison. Additionally, secondary data and primary data complete each other to enrich the empirical analysis. Finally, the research questions are answered and placed in a higher level framework that can be implied in other case examples.

In this study, there were two kind of interviews – theme based and semi-structured interviews. The first round of the interviews is considered as preliminary study. The second phase of the empirical study consisted of more specific questions to deepen particular sections.

## 4.3 First round - preliminary study data, data gathering and analysis

The main purpose of these first round interviews were to gather data of the case organization; to understand the operational environment, procurement as a function and specially, the P2P process. Similarly, there were interviews that gave insight on the technologies developed and utilized in case organization.

Altogether 15 different specialists and managers were interviewed (table 4). The participant list iterated during the process and was very wide. The snowball effect was utilized – the participants named a person who could be a suitable interview candidate. Most of the participants were from different sectors of procurement or IT. To point out, the procurement in this case organization is rather new in this particular form, so the experience among the participants in this specific company was at approximately three years.

For the major part of the interviews, the researcher's colleagues sent first introducing e-mails. After that, the researcher contacted the interviewees directly via e-mail. Approximately half of the interviews were more an open discussion within a specific theme – brainstorming around the thesis scope. The other half were more a learning session for the researcher to gather information about a certain process or technology.

All the interviews were conducted personally, except for one via phone. This was chosen to get additional value through facial expressions and gestures. Beforehand, themes were chosen depending on the specialty area of the interviewee, nonetheless the interviews were quite open discussions around the theme to gain a very wide perspective of the topics. These themes are roughly categorized into two. Equally importantly, during all of these interviews, notes were written and the first round of interviews were analyzed based on those notes. To emphasize, these interviews were not recorded. This way the participant was more honest and could concentrate on answering rather than holding information in the fear of judgement. All in all, these elements were deliberately decided to achieve the most straightforward and transparent answers. Correspondingly, a more honest relationship was built between the researcher and the interview participant for the further, second round interview.

***Table 4.*** *Preliminary interviews.*

| Organization | Responsibility | Theme | Date | Method |
|---|---|---|---|---|
| Procurement FI | Sourcing manager | knowledge engineering | 7.2.2017 | Theme interviews |
| Procurement | P2P-process owner | P2P | 13.2.2017 | Theme interviews |
| IT FI | RPA developer | knowledge engineering | 14.2.2017 | Theme interviews |
| Procurement | | P2P | 20.2.2017 | Theme interviews |
| Finance | Accounts payable specialist | P2P | 20.2.2017 | Theme interviews |
| Finance | RPA developer | knowledge engineering | 20.2.2017 | Theme interviews |
| Procurement | | P2P | 23.2.2017 | Theme interviews |
| Finance | Master data manager | knowledge engineering | 23.2.2017 | Theme interviews |
| Procurement | Process manager strategic sourcing | P2P | 23.2.2017 | Theme interviews |
| Procurement | | P2P | 23.2.2017 | Theme interviews |
| IT | Operational improvement manager | knowledge engineering | 3.3.2017 | Theme interviews |
| IT | Vendor manager | P2P | 6.3.2017 | Theme interviews |
| IT | IT architect | procurement data | 6.3.2017 | Theme interviews |
| IT | IT architect | knowledge engineering | 7.3.2017 | Theme interviews |
| IT | Supporting purchasing team leader | P2P | 10.3.2017 | Theme interviews |
| Procurement FI | Operational procurement coordinator | P2P | 9.2.2017, 13.2.2017 | Theme interviews |
| | | **Total** | **16** | |

Seven of the interviews took place in Stockholm and were conducted in English, whereas the rest of the interviews were in Finnish in Helsinki. Part of the interviewees were from the case company's group organization and part were from Finland's country organization. The interviews lasted from 30 minutes to 1 hour and 30 minutes, one hour being an average time. Two of these interviews were part of a colleague's introduction to the case company and were more of an overview of a specific procurement's function. Otherwise, interviews were conducted individually. In three of these interviews, the participant showed different systems or modeled new technologies for the researcher. Roughly in half of the interviews the participant had prepared a presentation beforehand for the researcher. Only one of the specialist was interviewed twice during this first round, due to the convenient location and the hands-on knowledge.

## 4.4 Second round - study data, data gathering and analysis

The goal of the actual empirical study was to gather information about the decision making in the case company and as well as the role of invoice data and PO data. In addition, there were questions about the advantages and weaknesses of the document analysis of the thesis.

Altogether 9 interviews, which are displayed in table 5, were conducted within the procurement and finance organization. The group of interviewees was deliberately considered beforehand and discussed with several persons to achieve the most suitable candidates. Only one of the participants was via recommendation of an interviewee. Most of the interviewee were high in the organization chart and were part of the leadership team. One of the interviewee was in the executive level. An important factor affecting the study was the major organizational change in the case company's procurement function. The organization was under turbulence and construction that complicated finding the right people in the correct position because roles were changed. Additionally, many interviewees in this section took part in building the organization so their time was limited at this particular time period. The interviews took approximately 1 hour.

***Table 5.*** *Empirical study interviews.*

| Organi-zation | Responsibility | Theme | Date | Method |
|---|---|---|---|---|
| Procure-ment | Chief Procurement Officer | Procurement's decision making | 30.5.2017 | Semi-structured interviews |
| Procure-ment | Head of Business Devel-opment | Procurement's decision making | 30.5.2017 | Semi-structured interviews |
| Finance | Head of Financial Services | Invoice data and decision making based on it | 30.5.2017 | Semi-structured interviews |
| Finance | Accounts payable man-ager | Invoice data | 30.5.2017 | Semi-structured interviews |
| Finance FI | Accounts payable special-ist | P2P process | 17.5.2017 | Semi-structured interviews |
| Finance FI | Accounts payable team leader | P2P | 23.5.2017 | Semi-structured interviews |
| Procure-ment | P2P Process owner | P2P | 30.5.2017 | Semi-structured interviews |
| | | **Total** | **7** | |

The roles of the interviewee and the interviewer were turned in the second phase. In these interviews, the researcher had prepared a slideshow of the thesis context and some results of the document analysis that were the foundation of the interview. This added the researcher's reliability and credibility.

The empirical study was conducted in from 17th to 30th of May. The invitations for the interviews were sent via e-mail that included a broader view of the thesis scope and example interview questions for the specific participant. The method used was semi-structured interviews to get answers to specific questions as well as create a holistic view of procurement. From these interviews 4 were placed in Stockholm and 3 in Helsinki. One of the interview was by phone due to unpredictable changes in scheduling, other ones were conducted personally.

## 4.5 Document analysis data, data gathering and analysis

The case organization has thousands of vendors in several categories so handling all these invoices or even a major part would be inefficient and take up a lot of time. This document

analysis focuses on invoices that have a PO. Overall, there were four different targets in the document analysis:

1. Analyzing the specific data requirements in the data sets
2. Enrichening the big picture with details
3. Categorizing the data by attributes
4. Conducting a comparison between invoice and PO data

The most central data in this research were the case organization's received invoices and POs. These invoices were retrieved from the case organization's ERP system and therefore, were confidential. The invoices were collected from the year 2016 to assure that also PO data would be available. All the invoices were paid and posted in the ERP system. Some vendors have only data from few months, if they had a sufficient amount of invoices during that period. Additionally, the data would be current enough. To clarify, there was no communication between vendors – only internal sources were utilized. The document analysis followed a particular step-by-step process that is shown in figure 13.

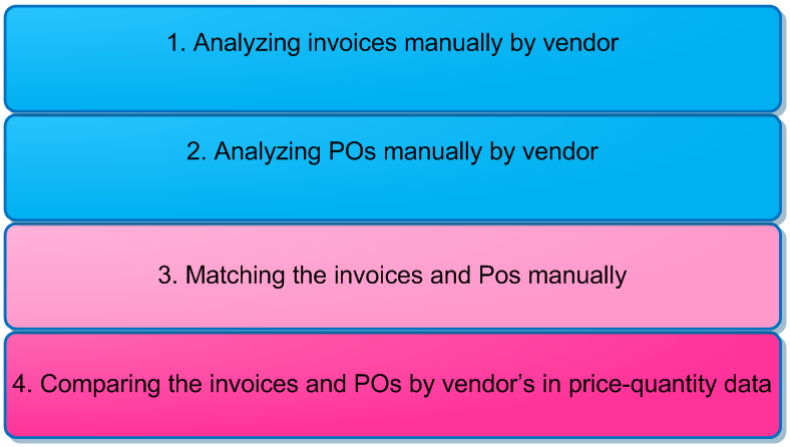



***Figure 13.*** *The document analysis structure.*

Altogether, 457 invoices were analyzed. The invoices itself varied from one page, with only one purchased service or product, to few page invoices with hundreds of different items. The paid amount fluctuated from ten euros to approximately million. A part of the invoices was in Finnish, but most of the invoices were in English. Even from the same vendor, the language varied. Most of the paid amounts were in euro, some in Swedish krona (SEK) and United States dollar (USD).

To point out, the selection of the vendors was a critical choice that determinate the whole document analysis. First, only two vendors were chosen but the invoice amount was too little. To increase the amount, there were two options: to increase the vendor amount or prolong the time period. However, from the invoice analysis from these two vendor's invoices it was quite clear that the quality does not differ much. The case organization's strategic and critical vendors were chosen because there is a lot of transactions and money

flow with these partners. These vendors also cover a range of different categories, providing an ambiguous perspective. In addition, the relationship with these vendors is collaborative, so based on the findings it would be possible to develop the invoices in the future. Of these critical and strategic vendors, the vendors that were known to produce good quality invoice data were eliminated because they have sufficient quality already. Altogether, 9 strategic and critical vendors were chosen.

In summary, the factors that determined the chosen invoices were vendor relationship and time. First, the invoices were retrieved from case organization's ERP system and these invoices were printed. The invoices were analyzed by vendor, comparing to the Pandit and Marmanis (2008) defined attributes. The invoice attributes were:

1. Vendor name
2. Invoice amount
3. Invoice date
4. Payment term
5. PO number
6. Due date
7. Item price
8. Discounts
9. Invoice number
10. Quantity
11. Item description

The analysis was conducted manually in Microsoft Excel by comparing are these attributes found in the invoices.

Gathering PO data was not as straightforward as collecting invoice data. In the first place, the PO data was also retrieved from the case organization's ERP system, first purchase order by purchase order. This means there were no wild invoices. These POs were printed manually and their length varied from a single page to 20 pages. However, these printed POs were only meant to be seen by vendors so many of the attributes were missing by default. Equally, the PO types varied a lot, so the data was different and fractioned in the PO data set. For an example, there are POs that are created in different systems or purchasing channels that leak to the ERP system. This made the analysis even harder as the data is not consistent. The robustness of the system challenged also the PO analysis as data is fractioned in many different data bases.

However, after analysis the researcher found out that a report from certain fields would be a more efficient way and it would decrease the probability of errors. These reports were pulled out from the system. The POs were also researched by comparing the data fields to the attributes Pandit and Marmanis (2008) described. The attributes were:

1. Vendor

2. Total purchase amount
3. Price
4. PO number
5. Cost center
6. GL code
7. Quantity
8. Item description

Pandit and Marmanis (2008) have also listed creation date, approval date and payment terms as important PO data but the reports did not include these data. Henceforth, these attributes are not included in the analysis. The reports also included a lot of data that was not in the scope of the analysis and first the data was cleansed for the analysis purpose. The actual analysis tool was also Microsoft Excel.

Last, the invoices and POs were compared. First, the invoices were mapped to the POs in order to study which data set has what data points. However, this mapping was aborted due to poor results that is explained in more detail in chapter 6.3. The invoices and POs were compared by vendor level in price-quantity data. This was conducted with the excel sheets generated before.

# 5. EMPIRICAL FINDINGS

## 5.1 The case organization's purchase-to-pay process

The P2P process is the core process of the procurement organization. The case organization's P2P process is depicted in figure 14 and the light blue boxes represent what the purchaser does himself. The challenging phases are marked by a red circle. The first step is to initiate the purchase request. In an organization that has several purchasing channels this can be a complicated task. The main purchasing channels are the ERP system and the web shop, which is a rather new channel in the case organization. The request depends on the chosen purchasing channel and it steers also the employees into different groups. Middle management in finance supports this statement:

*"This channel enables employees who have no clue to purchase, to buy stuff… the new purchasing channel has created a lot of problems."*

Comparatively, the web shop has also enabled employees, who do not have lot of knowledge about buying, to purchase in ease. The members in middle management as well as in the leadership team have stated regarding the web shop:

*"It has made buying much easier."*

There are employees who buy themselves but also employees that define their need and address them to procurement for advice. In this phase, procurement should be involved, regarding negotiating contracts with the vendor. Although this should be recommendable, the purchasers also negotiate with vendor and only ask procurement for signature. The executive level summarizes:

*"Horrible tools lead to not using the process and that leads to bad decisions. A regular person has already decided which vendor he is going to use before procurement."*

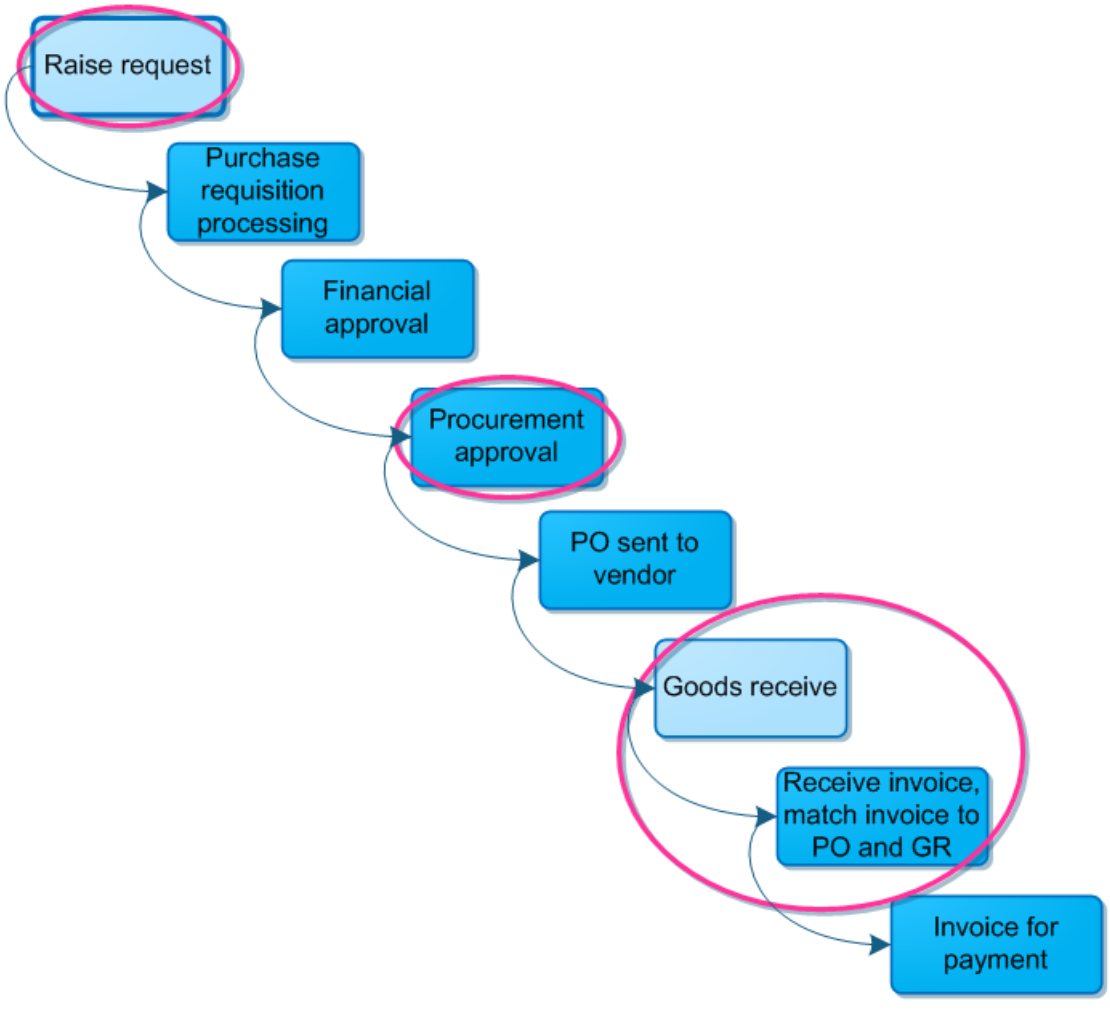



***Figure 14.*** *Case organization's P2P process.*

In the next step, purchase requisition process the PO is handled and checked if it meets all the requirements. This is only done for the web shop POs, for instance the right contract is matched to the specific PO. It could be stated, that this is a PO quality check point and it takes a lot of time as there are often errors. This means there is a severe gap in the POs that are created. Financial approval is an obligatory step where person assigned by the cost objective has to approve the PO. The next approval is from procurement, although procurement should have been in the process from the first step. From the researcher's point of view that this step is unnecessary and could be eliminated as process waste. POs are often sent from the purchaser's e-mail which is problematic as it does not increase transparent information sharing in the case organization.

When discussing matching goods receipt, invoice and PO the problems start to arise remarkably. It all starts with the goods receipt, which should be done after the commodity is received. Sometimes, the employees are not interested doing the goods receipt immediately when the goods are arriving. Particularly finance needs the information of goods receipt for accounting. The range of opinions is vast even in the middle management:

*"Most people just are interested that they get their stuff and don't do the GR."*

*"People do the GR after the invoice has come."*

*"From my perspective, doing GR is not important. The GR is only utilized that the finance has the right time period"*

Another barrier is for instance when considering IT services – should the purchaser rely on the vendor's estimation or wait for the hours to be fulfilled before doing goods receipt. Accounts payable does the matching of invoices and POs. This is often a challenge for accounts payable, as it is stated that only 2 % go through the matching process directly. A middle manager debates:

*"There is a problem with the agreement and the invoices not aligning."*

"*Accounts payable manually looks at the invoices and handles them. If the PO lines are missing, accounts payable adjusts the line. But is this really what procurement wants?"*

There are only few cases when the invoice is sent back to the vendor: if there is a wrong price, quantity or the invoice is lacking PO number or cost objective. On top of these matters, the ERP system is considered as rigid and it creates other problems. A manager explains a situation:

"*…IT function had a centralized team to merely support their purchasers and this improved their PO quality. In addition, one team had implemented their own systems to pass the ERP system's restrictions and illustrates the challenge."*

This example gives evidence that there are sub processes in the P2P process but they are not well communicated throughout the procurement.

## 5.2 The stakeholders in purchase-to-pay process

The stakeholders of the P2P process are the end users, sourcing and procurement and finance that is shown in figure 15.

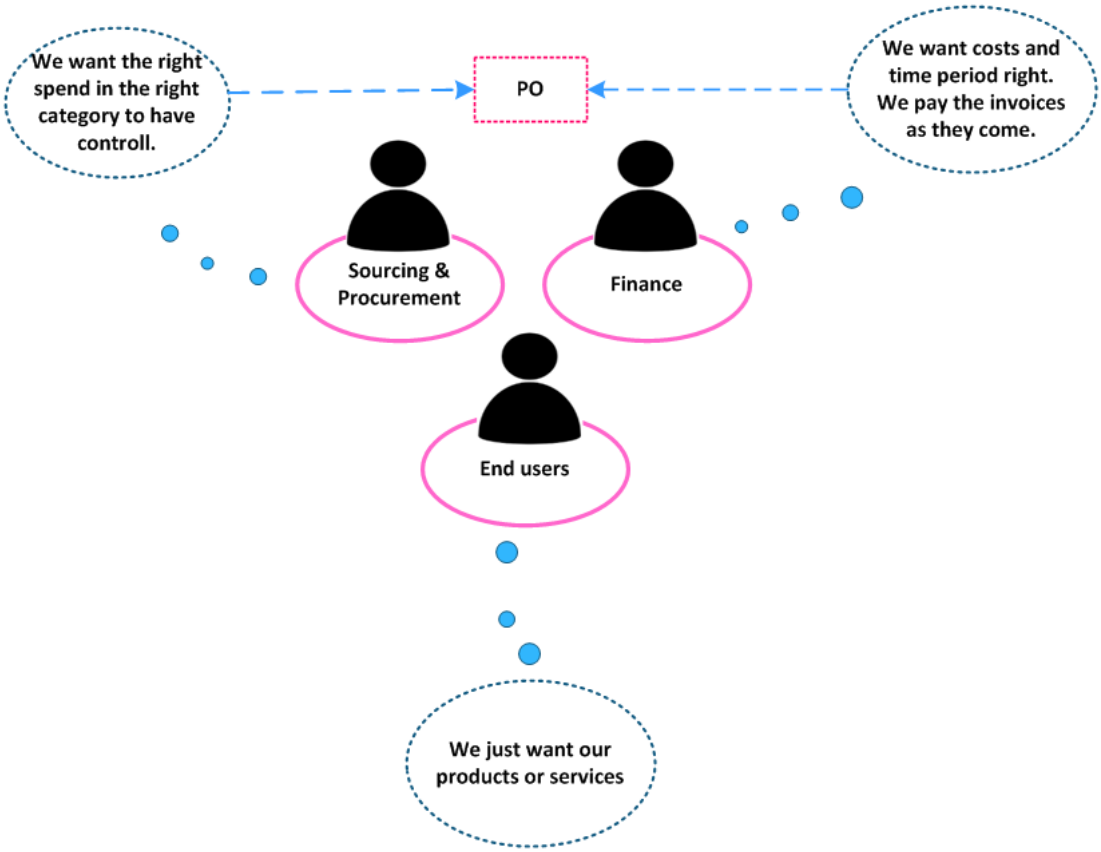



***Figure 15.*** *The different stakeholders in the P2P - process.*

The case organization clearly lacks collaboration between different functions. Even the functions that have a direct impact on each other, are not linked. The functions treat other organizations more from a static stakeholder perspective rather than functions that could co-create value. The stakeholders strive merely for their own target and lack of a united and shared vision. This chasm also reflects the data sets, as they are not completed with each other and not utilized efficiently. Besides, this impacts directly the reports and data analysis in the case organization, which leads to wrong or inaccurate knowledge sharing. Commonly in the case organization, procurement regards only spend that is in procurement scope – that is purchased by the sourcing and procurement function. However, finance looks the organization as a whole and has different numbers that does not synchronize with procurement. A leadership team member gives an example:

*"The wild invoices hold up 50 % of our invoices."*

At the same time, the managers in procurement argue:

*"The PO compliance is around 90% in most of the countries."*

The researcher asked for background information and validity about the numbers and it only caused more confusion and misunderstanding among some of the managers and specialists. In this case, this is an indicator that there is a severe issue within the shared data and reports. The PO is the link between finance and procurement which can be seen in the figure 15. This is elaborated in section 5.3.

The issue that the case organization is facing with the end user is straightforward: in the case organization everyone can purchase and this causes poor quality in the POs. However, it has been analyzed that in this particular case organization allowing every employee to purchase is more efficient than to have more procurement specialists. A manager in procurement states:

*"There was done an analysis that it is better to allow everybody to buy rather than have strategic buyers."*

A member in the executive team analyzes the problem of having lot of sourcing experts:

*"This is should be looked from the view point of scaling – at the moment the procurement personnel is overemployed but is not possible to get hundreds of people more to just purchase. It also depends what you are buying"*

Nevertheless, this should not mean that people are careless or do not follow the process. However, many interviewees stated that there is no strict policy and that employee behavior is one of the major reasons that is driving the poor PO quality. For instance, the case organization does not penalize the employees, who create poor POs or do not follow the process. They solely remind and guide the employee during the purchasing process. Moreover, the case organization supports the purchasers with trainings but all the employees do not participate. Members in the leadership team and middle management support each other strongly:

*"Mindsets need to be changed."*

*"We can do improvements… with the system and processes we cannot get very far"*

However, the executive level argues towards the complexity of the process that steers the end users towards lack of discipline:

*"Buying is so overwhelmingly complicated. Even I don't understand what information I should have and where the required information is found."*

As a result, it can be argued that the purchasers might want to follow the process but find it too challenging. For instance, employees do not want to dedicate more time to learn the purchasing process if they only need it once a year. It should also be kept in mind, that P2P process is not a particularly interesting process.

## 5.3 Current state of the data sets

### 5.3.1 Spend data

The document analysis results are divided into qualitative and quantitative results. The qualitative results are presented in the following chapters. The present state of spend data in the organization is highly dependent on which dataset is looked. Even though, the company group level has consolidated targets for the data and united practices, in different the countries have different spend data quality. Additionally, the quality regarding materials is better than services due to the complex characteristic of services.

Figure 16 illustrates the different dimensions of spend data in the case organization at the moment. The mandatory attributes are almost always found regarding each purchase – time, amount and the vendor. Secondly, the blue attributes reflect the dimensions that are sometimes challenging. If the purchaser data is found, the WBS or cost center is correct. WBS is work breakdown structure and in this thesis it means if there is a certain cost allocation for projects. This applies because WBS and cost center are linked to the purchaser strongly. The organization has multiple purchasing channels, so choosing the right channel is vital to link the right category. The most problematic issues lie in the transaction levels. Depicting the commodity is often ambiguous and flurry, sometimes even non-understandable or it does not exist in the worst scenario. This affects immediately the category choice and often misleads the user. The desired state of spend data would be as the figure 16, if all these mentioned attributes would fulfill the determined data quality standards. Instead of looking merely these dimensions, it must be mentioned that additional attributes exist and they would enrichen the data greatly.

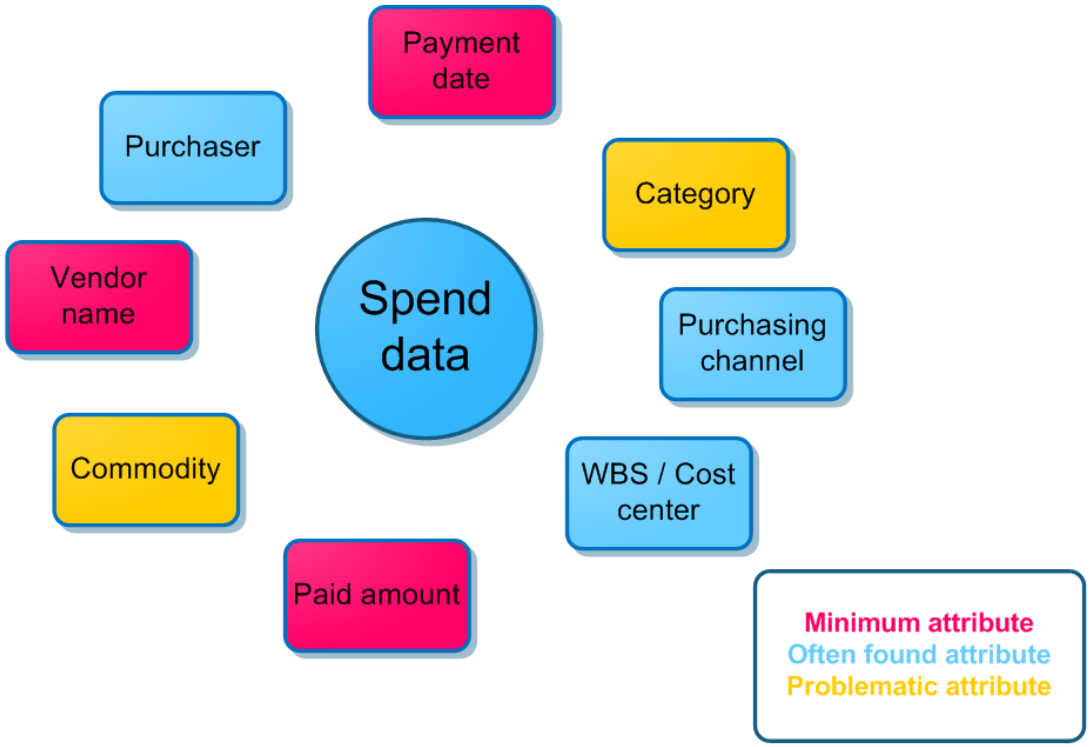



***Figure 16.*** *Spend data attributes in the case organization currently.*

### 5.3.2 Purchase order data

The case organization has the invoice data and PO data in the ERP system. This data flows into other systems to that are operational. Especially when analyzing the data architecture, the data source should be in high quality to assure that there is no errors in other systems too. To cleanse just one system would be inefficient as it is important to have the data correct everywhere.

Purchase order data is data, which is in the procurement scope. Most importantly, PO data depends directly on the purchaser – what is his capability of fulfilling the required data fields or his motivation of getting advice. There is a major difference between purchasing channels. The web shop has two different paths to buy: from catalogue or directly from contract. The catalogue web shop only requires four attributes: vendor, commodity and quantity, whereas the ERP system requires a vast amount of attributes. This results that there are attributes that are linked to the purchaser, such as cost center and organization by default. In addition, there are several attributes linked to the vendor and the product: material group, currency, tax and price. The contract option asks for the commodity name, price, unity and quantity. Nonetheless, the data from the web shop is leaked to ERP system. The ERP system requires a large amount of data inputs directly from the user, illustrated in figure 17. It can be concluded, that buying in the ERP system is more advanced and it demands more experienced users. A member in the leadership team argued that PO data is more accurate than invoice data:

*"It is more reliable cause its own our coding."*

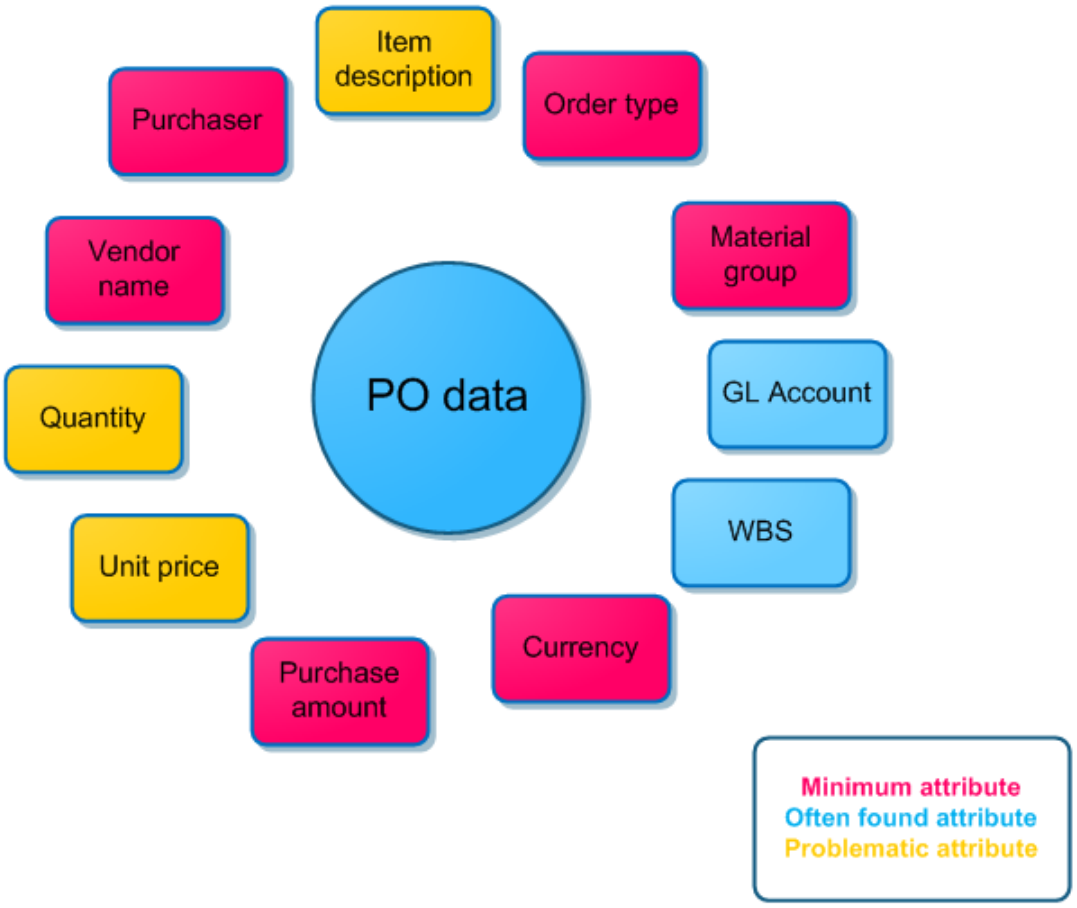



***Figure 17.*** *The current status of PO data in the case organization's ERP system.*

The purchase order data consists of 301 POs. The document analysis showed that order type, material group, purchaser, vendor name, purchased amount and currency were in all the POs. For example, without a purchaser name or vendor name, it is impossible to proceed in the ERP system. Equally, order type has to be identified and the purchase amount in the right currency. They are depicted as minimum attributes in figure 17. Although, the case organization claims that quantity and unit price are data, which without they cannot proceed, there are information gaps. Item description was in every PO, but because it is free text it is notably dependent of the purchaser. Additionally, GL accounts or WBS were found in almost every PO.

### 5.3.3 Invoice data

The invoice data is also from the operational systems that retrieve data from the ERP system. Invoice data is shadowed by poor follow up. For an example, the accounts paya-

ble does not have well-defined follow up as many tasks are handled manually. Nonetheless, there is a rigid process that is manual and has many exception queues, if there is information lacking. To put it in another way; there are severe issues within the invoice data. A member in the leadership team gives insight:

*"There is no proper follow up… accounts payable handles them manually and does not track them."*

The function for invoices is to be paid and it contains all the incoming invoices, not only in procurement scope. A member in the leadership team depicts two dimensions of invoice data:

*"Which vendor we have to pay… It is a legal requirement"*

There is a definite disagreement between finance and procurement people regarding invoice data reliability and quality. Finance illustrates:

*"Vendors cannot be relied. What if the price in the invoice is wrong for an example 110 when the contract says it should be 108?"*

*"Vendors do mistakes by purpose and by accident."*

The executive level in procurement depicts a way of handling the vendors:

*"Verify and trust"*

However, people consider invoices also as a reliable source because the invoices are paid and intentional mistakes cannot be conducted in the long run without no one noticing. A member in the leadership team summarizes:

*"The vendors can try to skate on thin ice."*

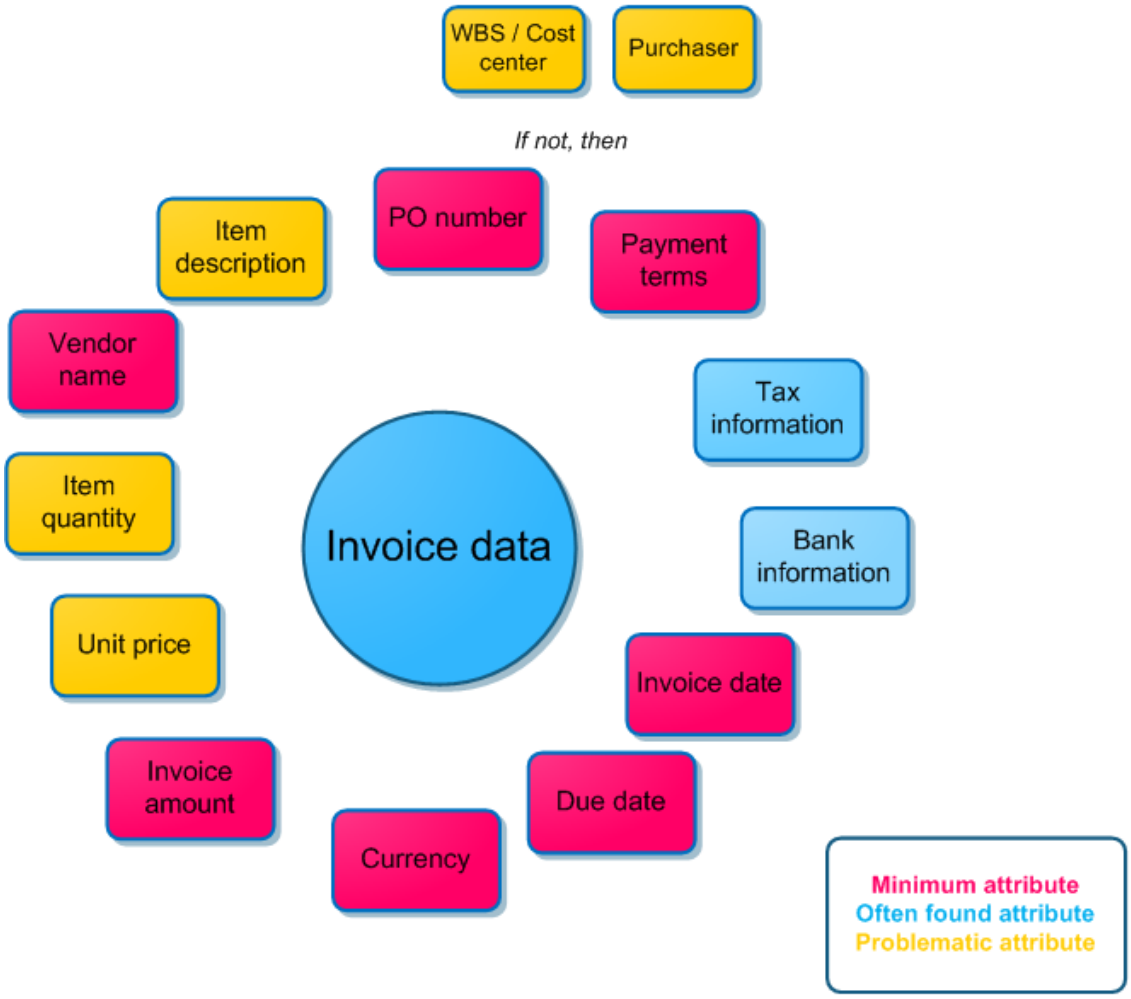



***Figure 18.*** *Elements of invoice data in the case organization.*

Figure 18 illustrates the dimensions of the invoice data currently in the case organization found out in the document analysis. All the pink boxes demonstrate the minimum data that is in this case vendor name, payment terms, invoice date, invoice amount, due date and currency. One vendor's invoices are missing due date. The PO number is also a mandatory element, however if there is a cost center or a purchaser the accounts payable can proceed with their process. As in PO data, also quantity and unit prices are problematic, though accounts payable has argued that those are obligatory elements. When looking at item descriptions, they are often coded with the vendor's serial or product numbers so they are hard to understand as a researcher. Tax and bank details are commonly written in the invoices.

## 5.4 Decision making in the case organization's procurement

Decision are made in several different levels – specialists, managers, leadership team members and executives make decisions every day. The information needs though vary

a lot. Figure 19 summarizes the main findings of decision making in the whole procurement organization. Procurement's main task is to decide whether procure or not, and if yes, then what commodity should be purchased in which purchasing channel. Procurement needs all the stakeholders' requirements to support the company, for instance in what to prioritize in purchasing. A member in the leadership states:

*"We fulfill the demand of the company to procure the services to our customers."*

Particularly, it is important to align with finance too, to ensure cash flow and the sufficient level of savings. Procurement also calls for information from purchasers, to approve the purchase requisition, to find the right vendor and negotiate with the vendor. Especially contracts often require tacit knowledge – knowledge that is shared from mouth-to-mouth imperceptibly in only specific groups. Throughout the interviews, spend data was regarded as an important data source; spend data is linked to many contexts. For instance, the spend implicates which vendor is a key vendor. A sourcing manager points out two decisions that are done based on spend data:

*"With spend data, you know where the company stands in a negotiation and how the cake can be divided… The spend depicts the category you are managing."*

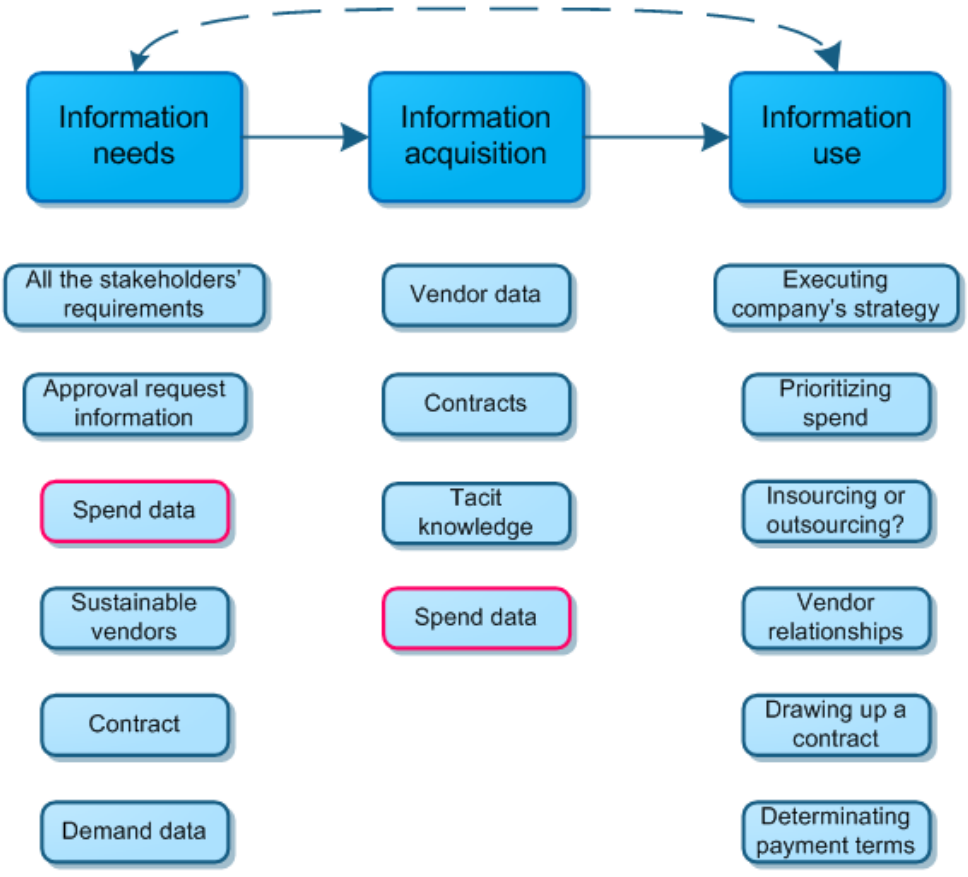



***Figure 19.*** *Information needs, acquisitions and use in the case organization.*

The information of sustainable and reliable vendors is critical and a member of the leadership team aligns with it:

*"Procurement ensures that the vendors are long time partners and this should be active decision... good relationships - not just monetary."*

Vendor data can be collected from contracts and spend data. Furthermore, benchmark data, reports and audits can be utilized to control the vendor relationships. The development of vendor relationships is regarded as top priorities in the case organization.

This can be looked from a broader perspective, such as: what are the company goals in two years and what investments should be done in a certain time period. Accordingly, there could be predictions for the future purchases. The decision of insourcing or outsourcing is a matter that encompasses also procurement. At the moment, there is no planning department that would provide the demand data for products and services. Nonetheless, it is acknowledged and there are plans for improvement.

In summary, there is a lot of data in the case organization but it might be hard to retrieve or to understand. The executive level summons it all up:

*"Data has to be used and it is an iterative process… The spend data will never be 100% correct. The main issue is how long it takes to fix things, not the issue that we have to fix things because that is inevitable."*

## 5.5 Summary of the impacts to spend data

The weaknesses are summarized in figure 20. The blue boxes illustrate the issues that arise from people – view; people are not caring and at the same the organization has no strict policy. The red ones reflect the process-view, for example the ambiguousness and the lack of systematic follow up or key performance indexes. Furthermore, the required information is hard to find and the system user-friendliness is low. The pink boxes demonstrate the challenges that stem from too many options of different matters such as purchasing channels, categories and vendors. The latter matter arose in many discussions and there was a uniformity of why the case organization needs so many vendors. Even still, the case organization practices spot buy to achieve better savings.

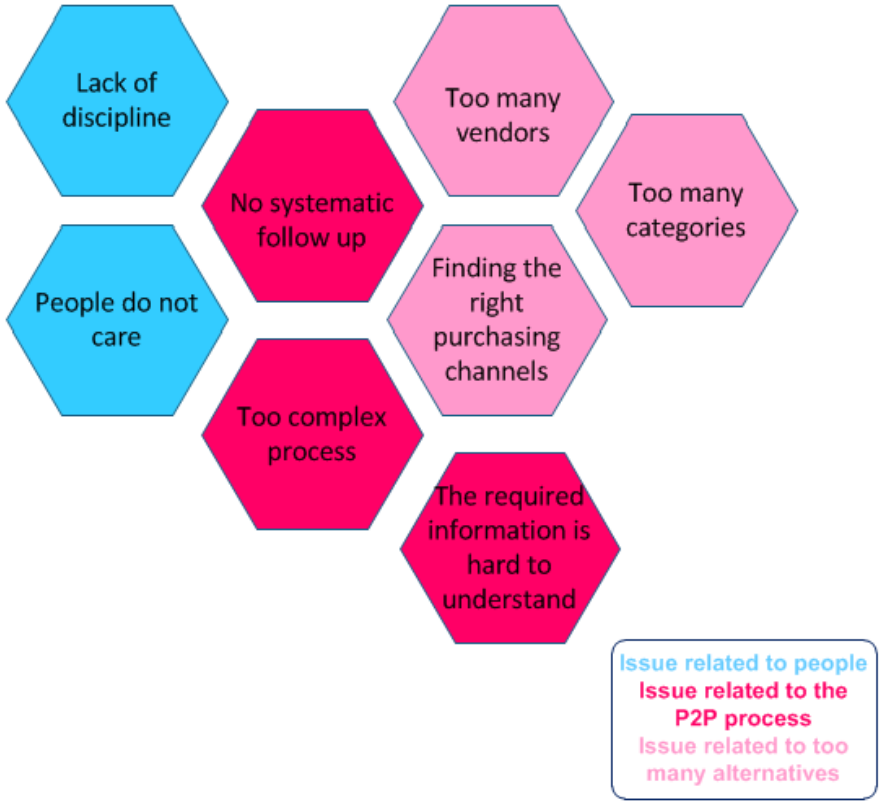



***Figure 20.*** *The roots of poor PO data.*

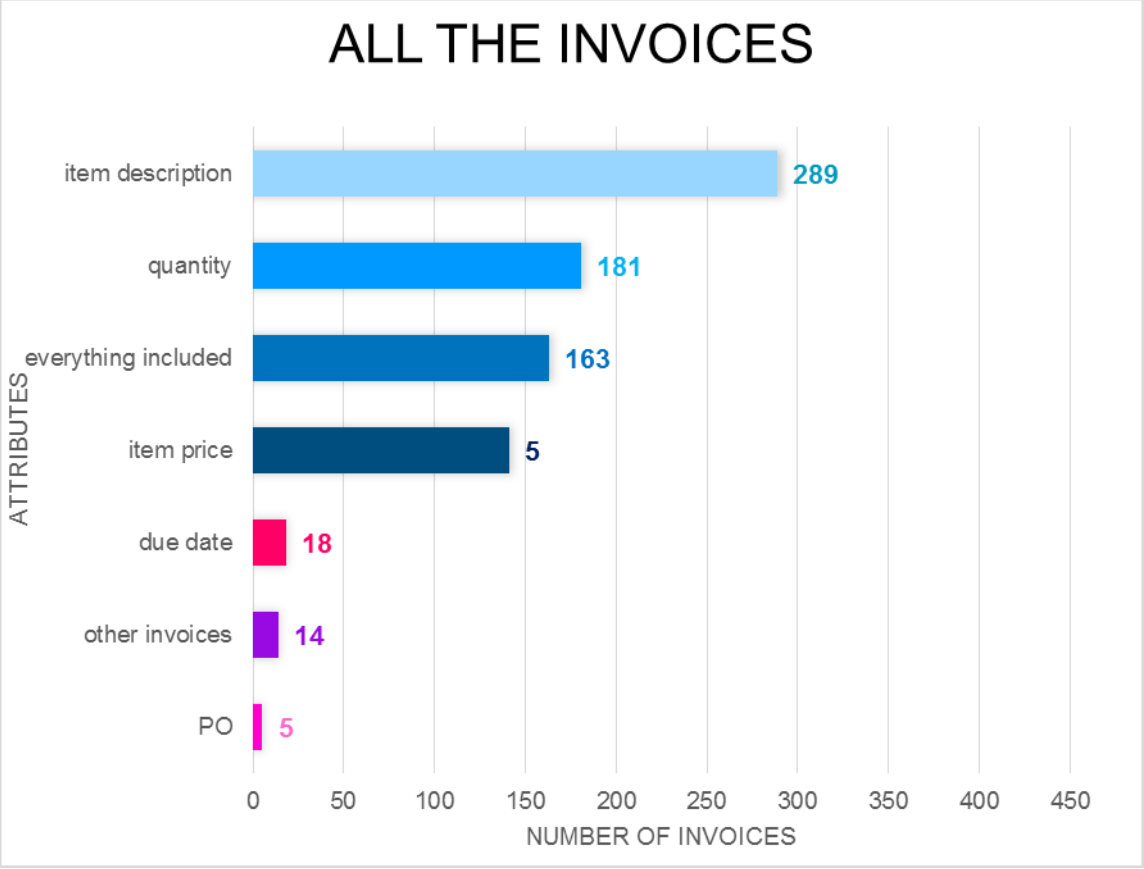
The ultimate strength of the case organization's P2P-process is that it is documented in an elaborated level. In addition, there is a system in place to create a PO and follow the process in a systematic way. Thereupon, it is way easier to build an improved process or to start making changes to the system to fulfill better the purposes. Everyone in the interviews were also highly acknowledged the problems, even though they highlight different themes. The top and middle management share the same ideas, so the strategy is well defined and communicated. It is important to realize that P2P is not a process that is the first one to be developed as it does not generate revenue.

# 6. COMPARING PURCHASE ORDER DATA AND INVOICE DATA

## 6.1 Invoice data set

These chapters describe the quantitative results and are explicitly based on the document analysis. Altogether, 457 invoices were analyzed. In this amount of invoices, there was a data saturation as the first smaller batch of invoices indicated the same results as the larger one. From this group 163 invoices had all the attributes. These were unevenly from many categories albeit IT service had the largest quantity and every IT service vendor was represented. Particularly, one IT hardware vendor had 61 complete invoices out of 74 invoices. In table 6, the whole data set is represented. In the invoice data set, there were also few credit notes that are shown as other invoices in the chart.

***Table 6.*** *Attributes that are missing in invoices.*




From all the invoices, 36% had every attribute. Mostly the invoices lack proper item description, being it in high percentage of 64. This might also be a consequence of the researcher's lacking knowledge of the commodities. However, the vendors slack creating a thorough item description as they might think that the purchaser knows what he is buying.

Second highest percentage is the attribute of quantity by a percentage of 40, this followed by item price a percentage of 31. This is further discussed in chapter 6.3. Often also item price is not included but it is tightly bounded to quantity. One vendor had invoices that miss due date, so there is a 0.04 % gap in due dates. On the positive side, the PO numbers were in almost every invoice, as only 0,01 % is missing. It can be stated, that invoices and PO's have a good link to each other. This is very important as it means that the data sets can be completed with each other.

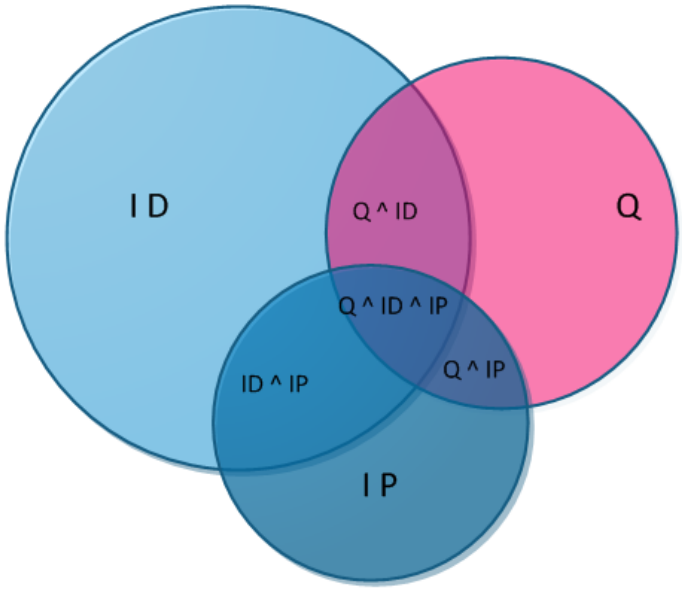



***Figure 21.*** *The relations between quantity (=Q), item description (=I D) and item price (=I P) in invoice data.*

In the analysis, there were several attributes that were lacking in the data with different combinations (figure 21). There were multiple different combinations. The main attributes missing were quantity, item description and item price in invoice data and therefore, they are chosen in the Venn diagram. It can be seen, that item description was missing the most, as the circle is the largest. Second place takes quantity. Quantity and item description were together lacking in 71 invoices. The combination of quantity, item description and item price occurs 89 times. Item price and item description were also a common combination as in 33 invoices they were missing. From the figure it can be seen that item description and item price lack also by themselves as quantity is more often lacking together with other attributes. Furthermore, it can be concluded that the attributes more stand-alone than occur in combinations.

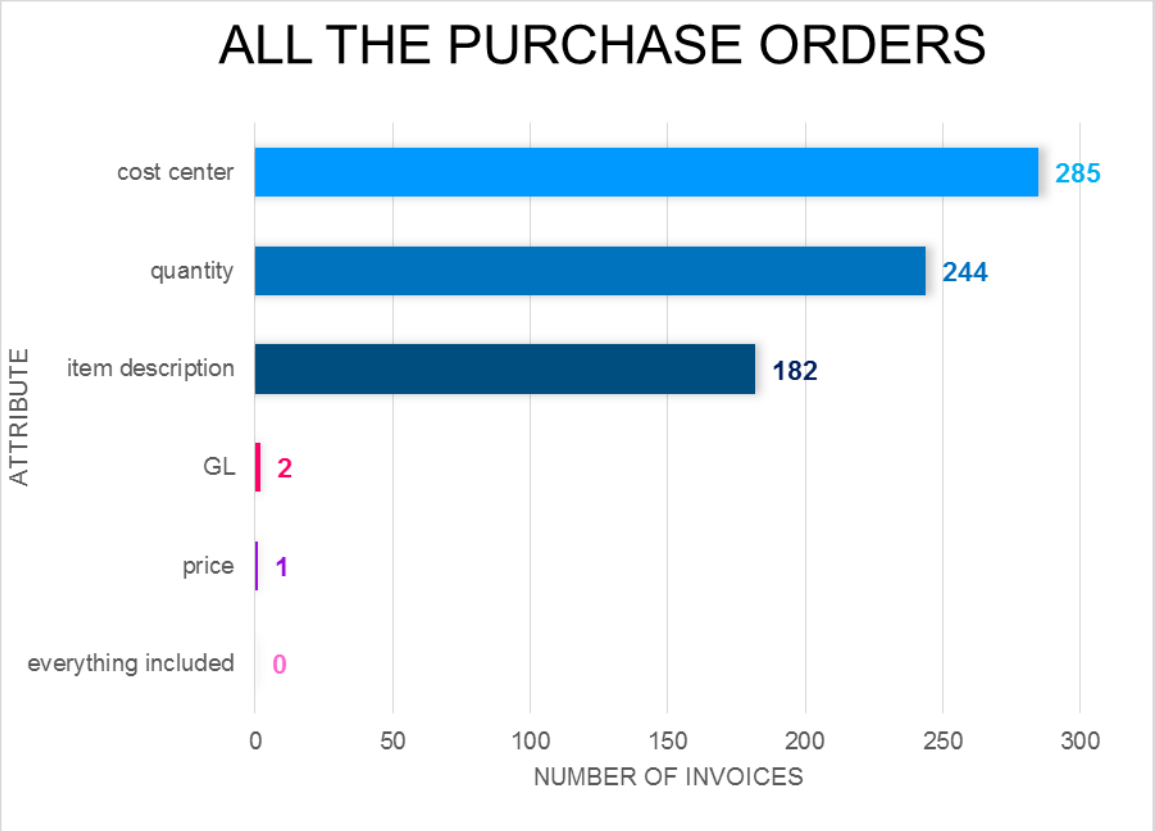
## 6.2 Purchase order data set

There was 301 POs to be analyzed that were linked to the invoices. There are a lot of variations in the PO's as some of the POs are created by expert purchasers and by regular users. The latter one is the minority in this research. Thereupon, as the thesis is focusing

only on strategic and tactical vendors, the PO's are created in a better level than usual. For an example, no maverick spend is accounted (such as office supplies or gifts) that is typically purchased via a robust segment of the employees.

First, because the POs were matched to invoices, the PO data set varies a lot. This means that the parameters were not altered to be suitable for the analysis, for example limiting the POs just to consider 2016. For instance, the POs might have also other purchased items than stated the invoices. To elaborate, there were POs that included commodities for several years. Another situation is that some of the POs had a large amount of money allocated but only a certain amount was used in that purchase. On the other hand, some invoices were larger than estimated so money was lent from another PO.

***Table 7.*** *Attributes that are missing in the PO data set.*




Altogether 301 POs were analyzed and from this batch none met all the attribute requirements. Table 7 reflects the missing data attributes. The major flaw is that 95 % of the POs miss cost center data, whereas they have GL data. However, every PO had a WBS so by definition they are not costs but rather long time investments so there are no cost centers. In addition, many invoices lack quantity, as 81 % of the invoices do not have quantity data. This is discussed further in chapter 6.3. Item description is also defined poorly in POs (by a percentage of 61), so this data point cannot be reached in neither invoices nor POs in a good standard. There were less than 0.005 % of POs that missed price or GL.

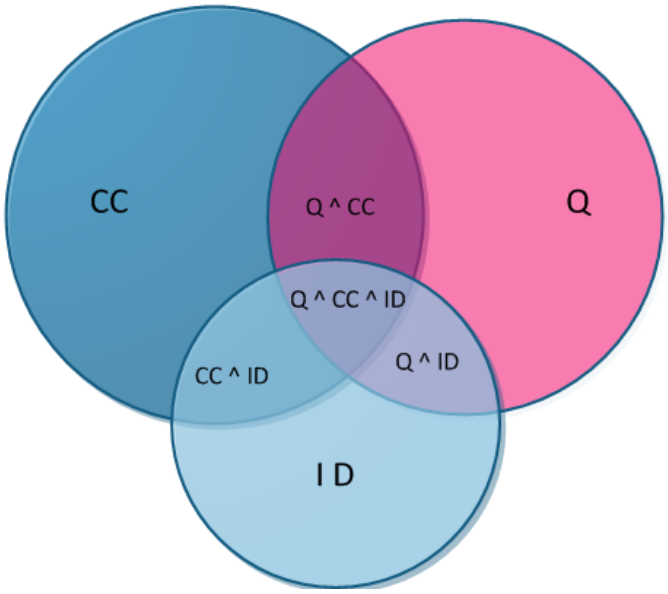



***Figure 22.*** *The relations between quantity (=Q), item description (=I D) and cost center (=C C) in PO data.*

The relations in PO data is quite balanced between whether they are linked or do they occur alone (figure 22). PO data lacks obviously most in quantity, cost center and item description. Quantity and cost center are missing the most together by 83 times. The combination of all these elements is represented in the PO data 150 times.

## 6.3 Comparison between purchase orders and invoices

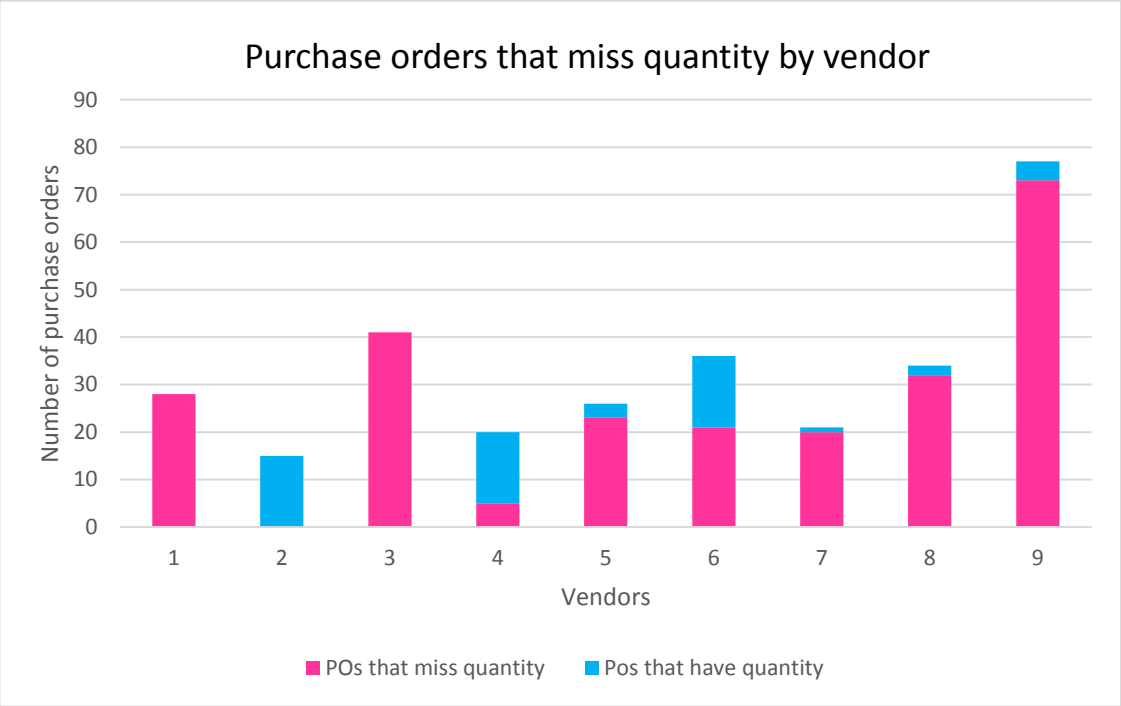
There were nine different vendors in three different categories to be analyzed. First, there was a manual mapping of the POs and invoices. There were four same attributes to be analyzed: PO number, total amount, item description and quantity. This was conducted by matching the invoice' PO numbers to the POs by vendor. In order to match the invoices, the POs should have been exactly the same as the invoices, first looking at the total amount. In this case they were not. When comparing the POs it turned out that the POs are rather estimations than precise numbers. This was particularly common regarding IT services. In addition, the POs might have only one line item, whereas there could be several invoices and vice versa. For this reason, the matching of invoices and POs was aborted.

By the same token, there was a problematic attribute regarding dates. There are several dates considering time, such as the date when the invoice is posted, the PO is created, the invoice is paid and the PO and the invoice is booked. Therefore, the dates were not compared limited out from the analysis.

## 6.4 Price-quantity data

The selected dimension to compare was price-quantity as there is a severe challenge in this data in the case organization. This attribute is unambiguous and can be defined clearly from the previously mentioned four attributes.

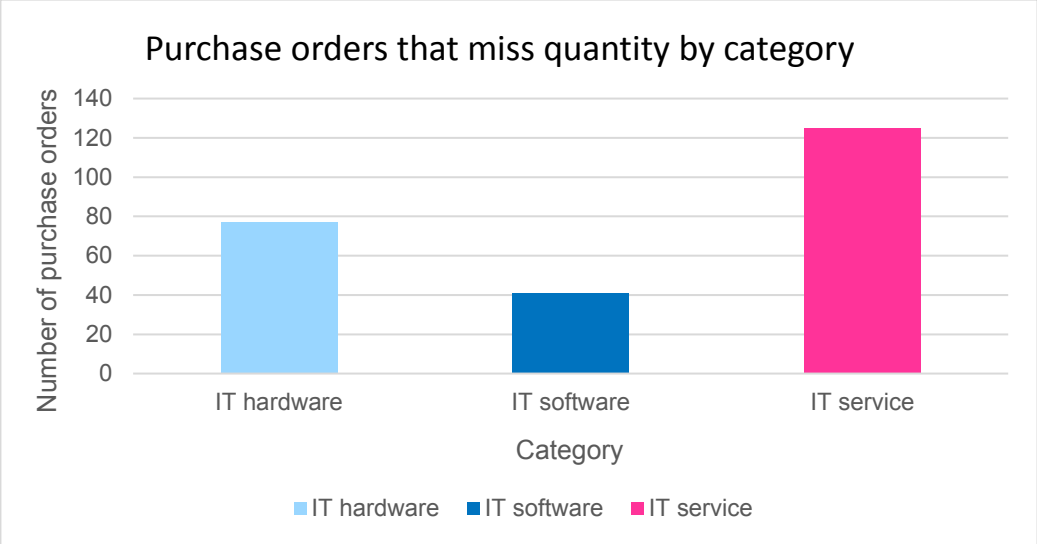
***Table 8.*** *POs that miss quantity.*




If looking at the POs in table 8, the bars POs that have quantity and the bars of POs that miss quantity are almost same in most of the numbered vendors. Six out of nine vendors' POs lack quantity in almost every PO. On the positive side, one vendor has PO's with quantity data. To point out, this vendor had commodities that are easily countable in this batch of POs. This can be translated that good understanding of what the purchased commodity is, leads to better quantity data. The remaining two other vendors in (4 & 6) have quite good quantity data.

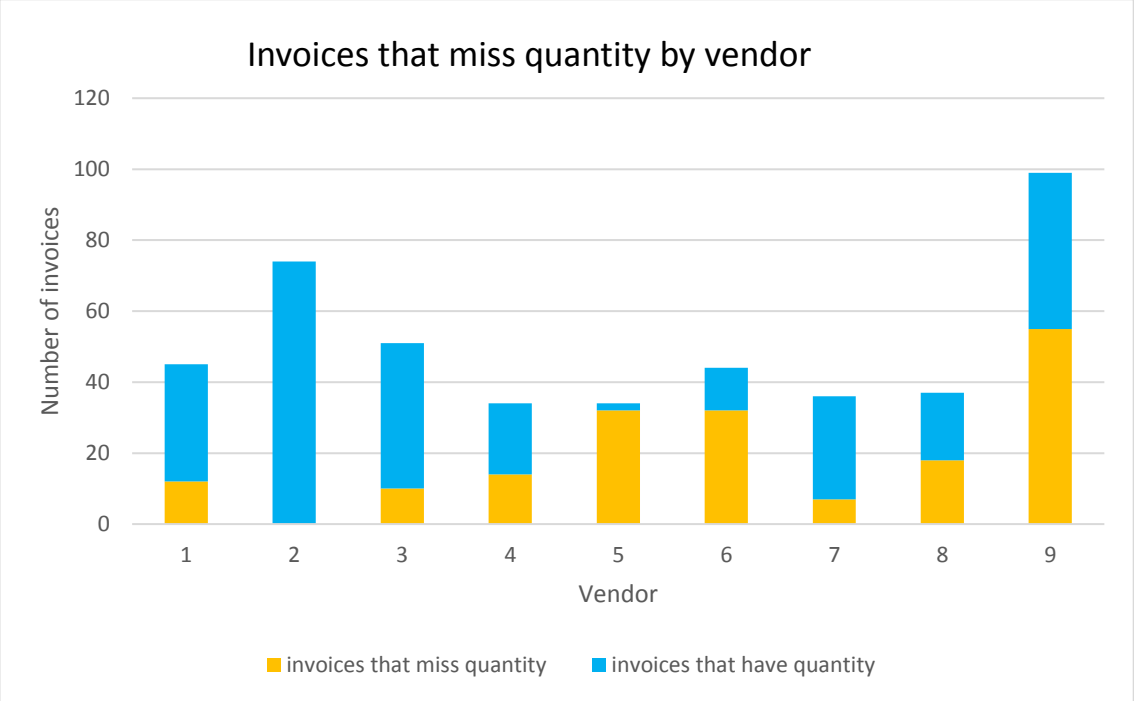
It is challenging to make defined reasons why the vendors' POs vary as there is no distinguished category that has better data quality. In all categories the POs are often submitted as having one (1) unit of the purchased commodity. Instead in real life there can be hundreds of different items or other invoiced commodities. These are sometimes mentioned specifically in the signed contract. Units also differ, are they time based or pieces. In table 9, the quantity is shown by categories: IT hardware has 5 vendors, IT software 1 vendor and IT service 3.

***Table 9.*** *POs that miss quantity by category.*



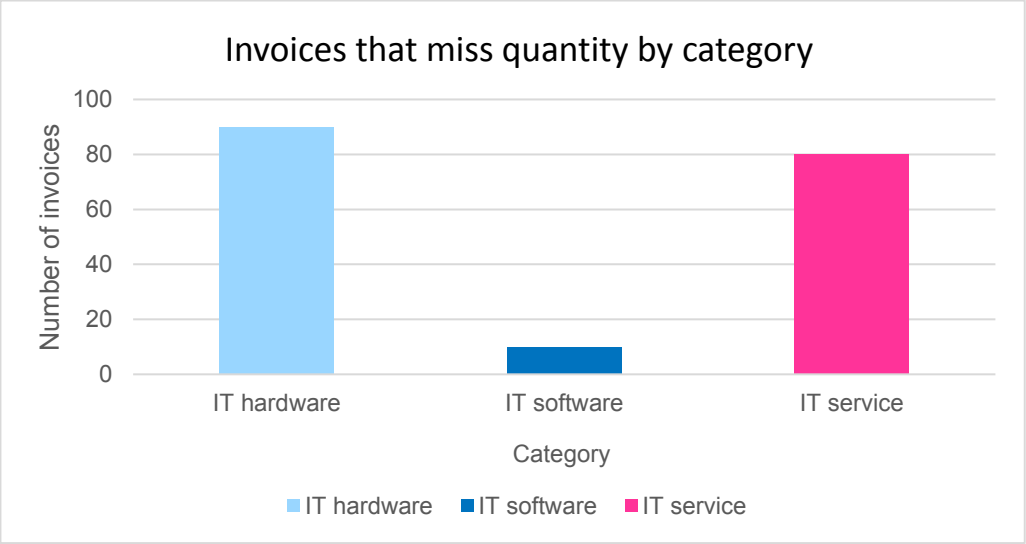

It can be seen, that IT service lacks quantity more often than the other vendors (table 9). For instance, buying services is a more complex process that leads poor quality regarding price-quality attribute. Services are mostly intangible and decentralized for a longer time period. Mostly the problem is that the unit of the quantity is not mentioned in services – is it months, hours, weeks or a contract work. On the other hand, vendors that provide many commodities are also problematic (IT hardware) – for a third party it is hard to understand what is bought in different cases due to the incomplete description. In this case, the offerings are quite large. For an example, a vendor under IT might provide IT software and IT hardware.

***Table 10.*** *Invoices that miss quantity.*



Regarding invoices (table 10) there are only three vendors (4, 8 & 9) that have almost exactly same bar in lacking the quantity and having quantity. Furthermore, there are three vendors (1, 3 & 7) that have only a minority lack in quantity in their invoices. Vendors number 5 and 6 miss quantity in most of the invoices. Surprisingly, vendor 2 has quantity in all the invoices.

***Table 11.*** *Invoices that miss quantity by category.*




In table 11, there is a representation of how the quantity divides between different categories. It is clear that IT hardware lacks the most, but IT service comes rather close. IT software has a minor share, however there is only one vendor in this category. Vendors make an effort to include quantity data in their invoices. At the same time, vendors also place sometimes one (1) unit with a large sum that consists of many commodities. Understanding invoices requires even more knowledge of the purchase as vendors have their own coding for the items. Finding the reasons why the invoices are not consistent would require surveying the vendors and tacit knowledge. For example, one interviewee stated that it is commonly know that one particular IT hardware vendor has always had poor comprehensibility.

If comparing invoices and POs, the hardware category reflects the main difference. In POs, the quantity is included way better. This might result from the fact that the invoices are coded and written by the vendors so the invoice data is harder to understand. Additionally, vendors have better quantity data in IT services than POs have. The most compelling evidence is, that POs are only roughly estimated for different purposes. They might be created for a larger sum of money. Equally, sometimes POs for IT service are estimated for a longer time period. These POs are just split to different invoices or even years. With this in mind, the invoices are usually for a shorter time period or project, so the quantity is shown more precisely. This is also the case for IT software. However, exceptionally the same IT hardware vendor has perfect quantity data in both POs and invoices.

The point overlooked is that these vendors are the strategic ones. The spend for them is major so the employees do not fix the POs to match the invoices exactly. Additionally, POs are made continuously for these vendors so if one PO lacks money, the money is taken from another PO for this particular vendor. Nevertheless, this distorts the spend analysis.

This case also reflects that the purchasers are quite sloppy so the policy in the case company is not rigid enough. On top of this, the interviewees have stated that making changes in the ERP system to correct the POs are very hard.

***Table 12.*** *The comparison between POs and invoice in quantity data.*

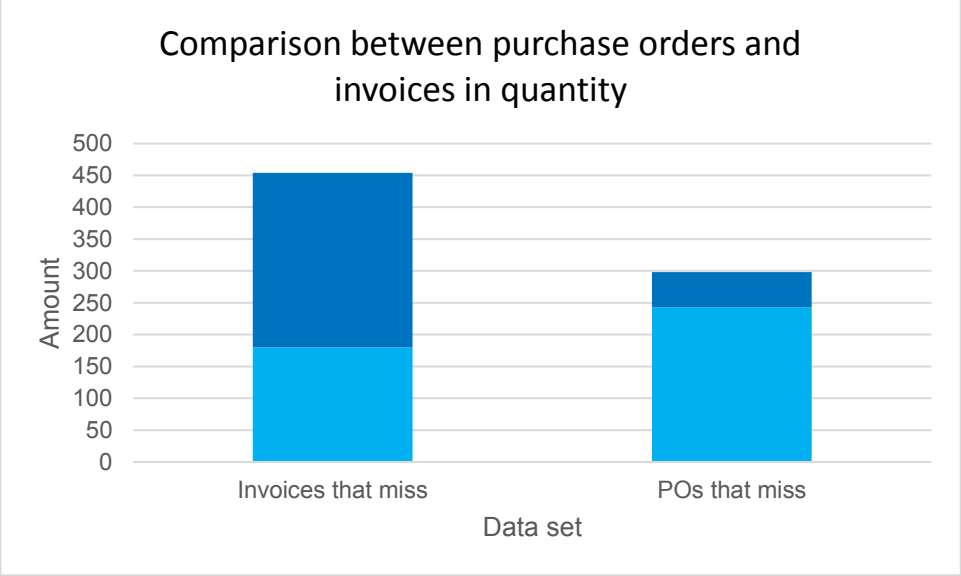



Table 12 shows the distinction between POs and invoices. In altogether 301 POs, 243 POs lack the quantity data. This means that 81.5 % of the POs lack quantity. Compared to the invoice set, only 180 from 457 invoices miss the quantity attribute. This leads to a percentage of 39.6 %. In this case, it is safe to say that invoices have better quantity data.

# 7. DISCUSSION AND CONCLUSION

## 7.1 Main findings

The first part of the first research question was ***what kind of decision making is conducted in procurement?*** Literature prioritizes that making the purchasing decision is the single most critical decision in procurement.

Procurement also has operational day-to-day decisions in the P2P process, such as deciding who buys, when and in which purchasing channel. The P2P process is an end-to-end process which aim is to purchase different commodities. It has six steps: specification development, gathering vendor information, vendor selection, drawing up the contract, the purchase ordering and the vendor evaluation. The first step, specification development is about determining the specific requirements for the purchase and it steers the whole process. After that, it is important to gather information about the vendor and the industry, for an example with references and recommendations. Vendor selection reflects choosing the vendor that meets the vendor criteria list best and it is one of the major decisions in purchasing. This determines the characters of the vendor-relationship; the strategical significance, the money and time spent, so it has a direct impact to the contract. Drawing up the contract has also many decisions such as time period, deliveries and terms of payment. The purchase ordering includes placing the order, receiving the goods and being invoiced. The last phase, vendor evaluation contains the assessment of both vendor relationships and the commodity. This navigates the decision is the organization going to continue working with the vendor.

To illustrate, there are fast and slow paced decisions that are both high and intermediate in importance. In the end, it all comes to serving the organization's high-level strategy. This means that procurement prioritizes what to purchase and when. It also takes a stance in outsourcing situations, to find the best savings.

The second part of the first question was ***what kind of information needs are there?*** Literature only scratches the surface; it depicts that contracts need a lot information and the ultimate decision to buy or not is also heavily information-dependent. Above all, these information needs are very much organization-driven. Van Weele (2005, p. 40) highlights an interesting point – procurement's gatekeepers that disable the transparent information flow. In reality, often the purchaser is not the user of the commodity and therefore, all the operational information needs might not be met. The thesis presents the relationships for identifying information needs and desires. There is information that is desired but not needed, which could be in procurement. Also there is information that is both desired and needed, for an example the payment terms for a contract. Information that is needed but not requested could be the risk assessment of the vendor to find out the due diligence.

Above all, there are information needs that are not yet identified as a desire or need. T The most important information needs for procurement are demand data, sustainable vendor and spend data, not to mention the importance of aligning with the stakeholders throughout the company.

The third question is regarding quality in three perspectives: ***What is the quality of purchase order data, invoice data and spend data in the case company?*** These data sets are mostly generated in the P2P process but they are enriched with various other data sets, which affects the data quality. Purchase order data is created when placing the purchase order for the vendor in the organization's own system. The data set should contain vendor name, creation date, payment terms, PO number, approval date, cost center, GL code, quantity, item description and total purchase amount. Invoice data on the other hand, is produced from the actual invoice the vendor sent. Invoice data has several attributes: vendor name, invoice amount, invoice date, payment terms, PO number, due date, item price, taxes, surcharges, discounts, invoice number, quantity, item description and total amount billed. Spend data is the combination of PO data and invoice data. The most important spend data attributes are the payment date, paid amount, vendor name, commodity, cost center and organization.

Data quality in spend management is mainly the accuracy of spend data. Spend data has challenges in data quality, as it depends on the P2P process. However, the data is also produced in other contexts and IT systems that have a negative impact as the data is not consistent and has to be normalized for spend management purpose. Data quality has various attributes, but in this thesis the dimensions for spend data are accuracy, reliability, timeliness, completeness and relevancy. Accuracy is that the data is precise and correct. Reliability refers that the data can be relied upon and it reflects the desired data. Up-to-date data with good accessibility at the right time is considered as timeliness. Completeness is data, where all the necessary attributes and entities are found in the data. The applicability and usability of the data shows that the data is relevant.

From the literature's (Pandit & Marmanis 2008, p. 5) 11 mentioned data attributes, mostly three attributes (cost center, quantity and item description) were missing most commonly. However, even though cost center was missing the case organization the same information was hold in WBS so the importance of missing cost center is low. Item descriptions were very high-level or heavily context-related so understanding those as a third party is challenging. The invoice data quality was significantly better. In invoice data, from the 14 different literature specified attributes also three attributes were missing mostly. These were item price, quantity and item description.

Invoice data is often utilized for spend data. It is surprising that holding the most basic spend data elements – quantity and price, is difficult to achieve, though it comprises the base for spend data. The research compared the price-quantity data as it was a dimensions that was interchangeable. In the empirical study, the research showed that invoices have

more complete data sets. Therefore, this thesis suggests for the case organization to enrich the spend data with even more invoice data.

Together these questions form the main question that is ***how procurement's decision making can be improved with enriched data.*** The decision making would be more accurate and data-driven with improved data quality and as a result the decisions would be better. When data quality is good, people would also rely on the data and consequently, utilize the data more for various decision making situations and create a positive loop of data iteration. The data would produce a framework to make several typical decisions in procurement and could be enriched with tacit knowledge to produce valuable information.

It is possible to see the patterns and trends of different purchases historically and on different segments, such as country-level, spend amount and category. This would have a direct impact for improvements in many areas, such as prioritizing spend, creating savings, utilizing compliance vendors and contract management. The organization can deep dive into specific questions, for an example to find out the reasons why a certain commodity is having high costs or why purchasers are using no-contract vendors. This means that the organization can improve operational performance and taking a leap towards strategic decisions. Taking control of these issues, the management of spend data improves and it can lead to better level due diligence and consolidation. In addition, there can be questions asked regarding wild invoices and off-contract spend: "Why is this happening? Are these vendors giving lower prices or better service level?" Without quality data, it is hard to prove these issues and make a change in decision making. On the whole, the decisions would support the organization's strategy.

## 7.2 Theoretical and practical implications

### 7.2.1 The P2P process

The case organization's P2P process differs mostly from the represented purchasing model, mainly adapted from Van Weele (2001). As the P2P process is characterized by operational elements, it was known beforehand that the theory framework will not fit precisely the P2P process. If comparing these two models, the major deviations stems from the strategic level. The figure 23 illustrates the differences. The blue represents the case organization's P2P process and the pink ones the theoretical purchasing model.

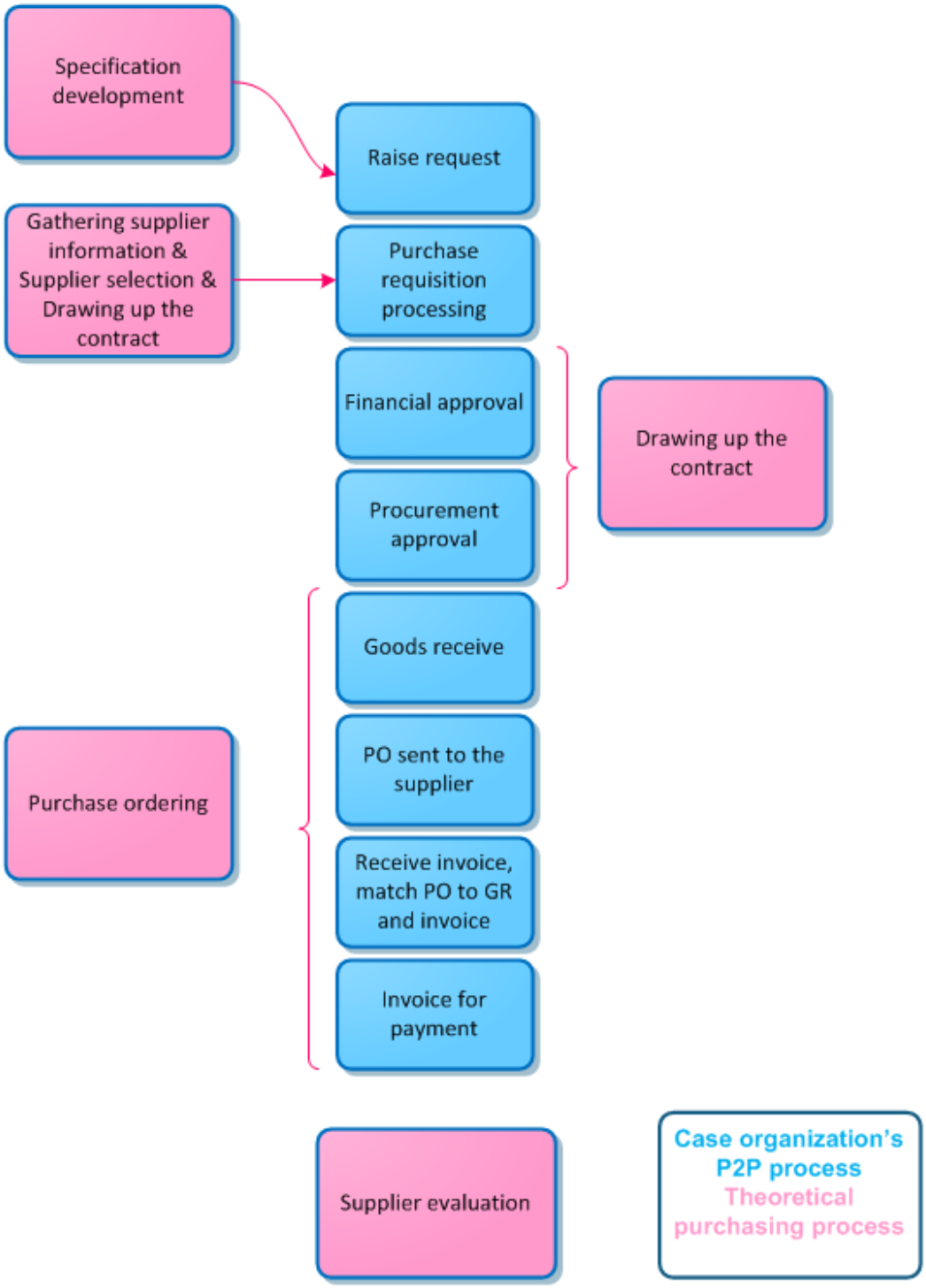



***Figure 23.*** *Comparison between the theoretical purchasing model and the case organization's P2P process.*

It shows that in purchase requisition processing there are many steps from theory (gathering vendor information, vendor selection and drawing up the contract), and on the other hand the two different approval are also a part of drawing up the contract. The purchase ordering is depicted in the case organization's P2P process in detail but the process lacks

the vendor evaluation. Summing up, the theoretical purchasing model has a different approach and more weight on strategic purchasing, but the operational P2P process supports it.

### 7.2.2 The spend data quality

Figure 24 compares the theoretical perspective and what was found in the empirical research. The red boxes are the ones that are mentioned only in theory and the blue ones represent what were results in empirical research. There are few exceptions that the researcher made. Instead of separating organization and geography, which was listed in theory and purchaser, which was mentioned in empirical research, they are combined as they represent the same data in different levels. In some cases, knowing exactly who purchased might be vital – for an example if the invoice or goods are wrong. Often, there are backlog data for the purchaser, such as location, organization and cost center. Therefore, they are linked together. In addition, either one attribute is in spend data: cost center or WBS and they provide the same information about allocating the money. On the other hand, commodity and category could be integrated as they also implicate the same data, only on different levels: what was bought. GL code is an element that should be examined as it could be potentially utilized for categorization.

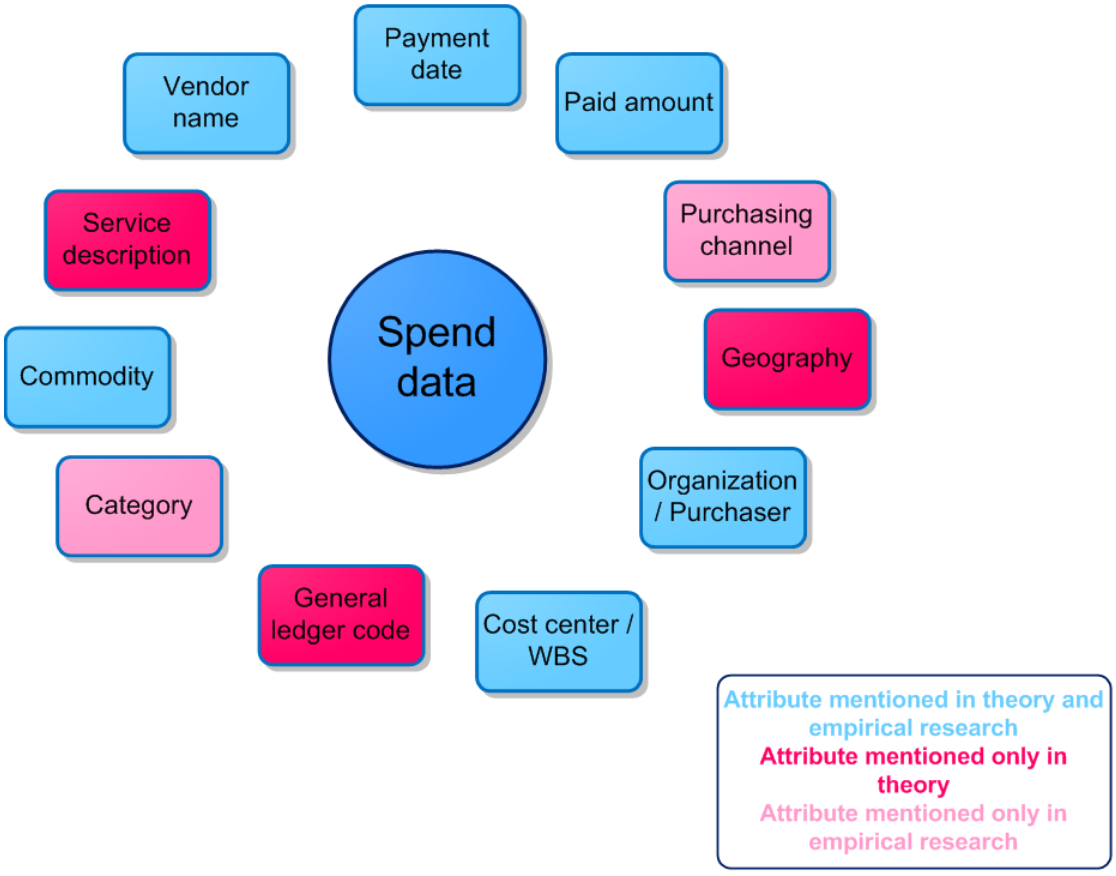



***Figure 24.*** *Comparing the spend data elements in theory and empirical research.*

If compared what was mandatory in Pandit and Marmanis (2008) framework for spend data, all the mandatory fields were found in the case organization's data: payment date and paid amount. Also the highly advisable attributes, organization, commodity, cost center and vendor name were in the case organization's spend data. From the advisable variables, service description and general ledger code were not always found. One main finding was that theory emphasizes significance of item description attribute but in the case organization it was held as high. This is quite understandable from the procurement side, though it would provide better data for spend analysis by category or commodity. From a practical point of view, it is a hard task to improve the item description as it is free text. Besides, this could be avoided if there was multiple selection that guided the item selection – implementing a tree for instance. For an example if the purchaser needs IT, first there could be a question for the most common ones, such is it a laptop or a router. However, from the tacit knowledge point of view, sharing information would be more transparent with accurate and precise item descriptions. All in all, it can be concluded, that the theoretical framework matches the empirical research results relevantly well.

The spend data quality framework in this thesis is demonstrated in figure 25 with combined in which part of the empirical research these attributes were researched. It can be seen, that researching accuracy in PO and invoice data set was not concluded as the matching was aborted. This means that were no comparison is the data correct and as the researcher did not obtain expertise of the subject, it is very challenging to say if the data was correct or precise. The assessment of reliability is highly dependent if the researcher managed to involve the right people that have the right data access for POs. For invoices, the data was reliable due that it was the direct invoices from the vendors. Because this thesis was written for a company as an assignment for their purposes, it could be claimed that the data is the desired one. The relevancy of the data was the main subject in the interviews and it was covered, nonetheless it relates directly how the researcher has understood the factors.

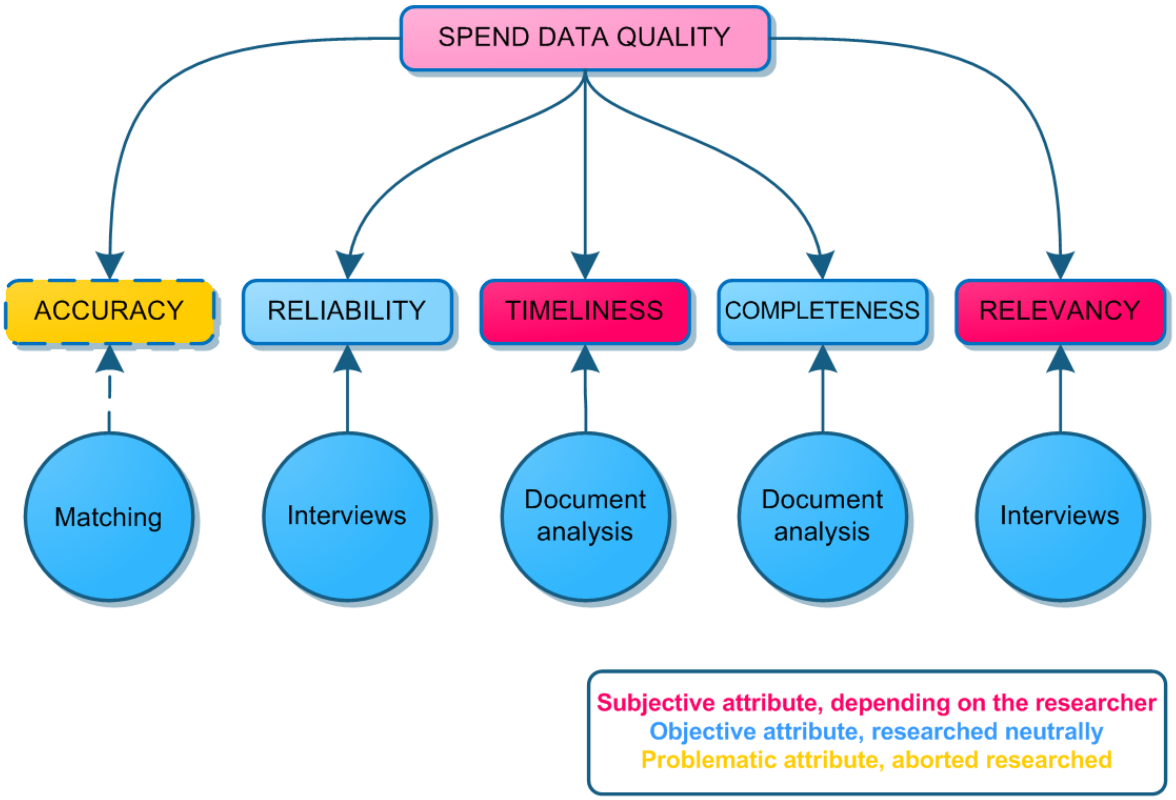



***Figure 25.*** *Concept of spend data quality.*

Many interviewees discussed that the relevancy is low from their point of view but if the relevancy would be poor, there would be no good decisions or decisions at all. Therefore, the researcher would clarify that the relevancy depends which decision and data set is looked. On a deeper level, the applicability for spend data is spend analytics and that still being in its infants in the case organization, tells itself that the applicability is on mediocre level. Timeliness was also an attribute that was part of the interviews. Completeness was the data quality dimension that was thoroughly researched in the document analysis. The document analysis's main purpose was to examine are the Pandit and Marmanis (2008) identified data attributes in the data and this was also precisely explained in the study.

## 7.3 Choosing purchase order or invoice data set for spend data

The invoice and PO data formed the foundation for the whole study. They have different characteristics that are listed below in the table 13. The main deviation between these data sets is the fact that purchase order is the organization's own and invoice data is from a third party. This means that the purchase order data can be controlled, collected and coded the way the organization wants and needs. Especially for spend data, the data coding is significant. Purchase order data usually is collected from ERP system but there are various alternatives, such as if there is a web shop for purchasing. Organizations adjust their ERP system to fit the industry and organization preferences so they differ a lot. Purposes for producing and collecting PO data vary as some organizations have them merely for

the vendor and some might have for historical data and to even forecast the future purchases. Although, invoices come more often from a non-stable environment there are legal requirements for the data so they are more consistent.

***Table 13.*** *Differences between purchase order and invoice data.*

| | Purchase order data | Invoice data |
|---|---|---|
| ***Ownership*** | Organization itself | Third party, a vendor |
| ***Data source*** | Usually from ERP system | An invoicing system |
| ***Purpose*** | Send information for the vendor and collect data for internal usage | Bill the organization for services and goods and obey the law |
| ***Data structure*** | Includes several data fields that are missing in the invoice data | More information about the vendor |
| ***Advantages*** | The coding and collecting of the data is for the organization to be decided | Legal requirement forces invoice data to be relevantly reliable |
| ***Disadvantages*** | Not so critical data base for most of the organization that causes poor behavior among the members regarding data quality. | Data from a third party, various vendors |

Both of the data sets are not meant for spend data and this causes challenges as the data has to be cleansed and transformed before utilizing as spend data. For the invoice data, the work load is greater to make all the vendors' invoices in the same data format because PO might come from one or a couple of internal systems. In addition, both data sets suffer from a spectrum of different users from beginners to experts.

It has to be also remembered, that PO data is first produced and based on it the invoice data is created. Moreover, there are some data fields that are not included in invoice data such as category, general ledge code and cost center. The problem stems from the fact that the vendor creates the invoice data based on other data sources than PO data, for instance members of the organization or personal e-mails. This causes disruption in the data sets and complicates the communication between stakeholders and documentation.

## 7.4 Recommendations for the case organization

The case organization should strive for solutions for the previously analyzed challenges. These were illustrated in figure 19 and they are listed below:

1. Lack of discipline
2. People do not care
3. Too complex process

4. The required information is hard to understand
5. No systematic follow up
6. Too many vendors
7. Too many categories
8. Finding the right purchasing channels

All data stems from the P2P process. Lack of discipline stems from both poor policies as well as poor organizational management. On top of this, the policies are not the transparently communicated to the employees. The careless mentality combined with the lack of discipline ends up with a major issue within the human factors.

First, policies and practices should be simple, meaningful and easy to retrieve. Besides, these policies should be transparently communicated to the employee – one method is to create a mandatory online training. The training could include information about the P2P process, most common exceptions in purchasing and who to contact if needed. In addition, the training could hold information about the severity of violating the policies. After a certain amount of reminders of the training, a supervisor could require the employee to complete the training, adding also discipline into the organization. Although there are employees that have tasks to help other employees to purchase, to find out the right channel is hard. People get frustrated, if they have to spend a lot of energy to resolve a relevantly insignificant problem. Therefore, creating one transparent is important. In this hub, information about which purchasing channel to use for different occasions should be also provided. Nonetheless, the purchase channels should be identified for different purposes beforehand. The aim is that a person from the organization could be able to answer where he could ask for help regarding purchasing – clear organizational management. After that mindsets could change more easily; changing mindsets is highly dependable in transformation management, the employees themselves and leadership. This thesis does not take a stance on those issues, however the researcher recommends the case organization to analyze the issues that are more on the soft side too to understand why employees are not obeying the P2P process. After that, it would be possible to eliminate process waste in the P2P process.

Simplifying the process means that the policies should be simplified also. With simplified policies and guidelines, the required information would be easier to understand throughout the organization. For instance, there are lot of good policies regarding due diligence as it is held as one of the most critical policy. The procurement organization is quite vast and spread out to also other departments, so building a common ground is an important step. Moreover, the case organization should start building bonds with the different stakeholders and unite the goals; not have separate missions for every department even though they are part of the same P2P process. This could be executed via common meetings where matters and opinions about the subject could be exchanged. The researcher would suggest creating a core team where a few employees of the different stakeholder groups would represent their own group to create common practices. This been said, the policies,

practices and the P2P process should be described in a manner that employees outside the procurement organization can easily understand. In addition, with enhanced collaboration, the P2P process can be simplified with fewer check points if these matters were discussed beforehand. This would also reduce the misunderstandings and work load considering correcting mistakes in the POs.

For improving the systematic follow up, there should be also a policy that includes documentation and data collection. If there is a lot of manual work, it makes the following up more difficult. However, the case organization has a lot of data that just has not been utilized – the information should be acknowledged and easier to retrieve. Nonetheless, building up dashboards and analytics that directly show patterns is the second step, although, people should understand why the data collection is important and without a demonstration it might be not high on their priority work task. Thereupon, showing the impacts of following up different measurements is the key.

The organization has already taken steps to reduce the vendor and category amounts. The contract data base has been analyzed to find out are there vendors that are not used. Creating a short list for the most preferable vendors might reduce the use of too many vendors. To change the categories, the category tree should be modified – this is not an easy task as for instance, all the accounts are mapped to the category tree. The figure 26 sums up all the recommendations for the case organization regarding the mentioned challenges. The blue color reflects the organizational problems, the red boxes represent the issues the lie within the P2P process and the pink one is about too many options.

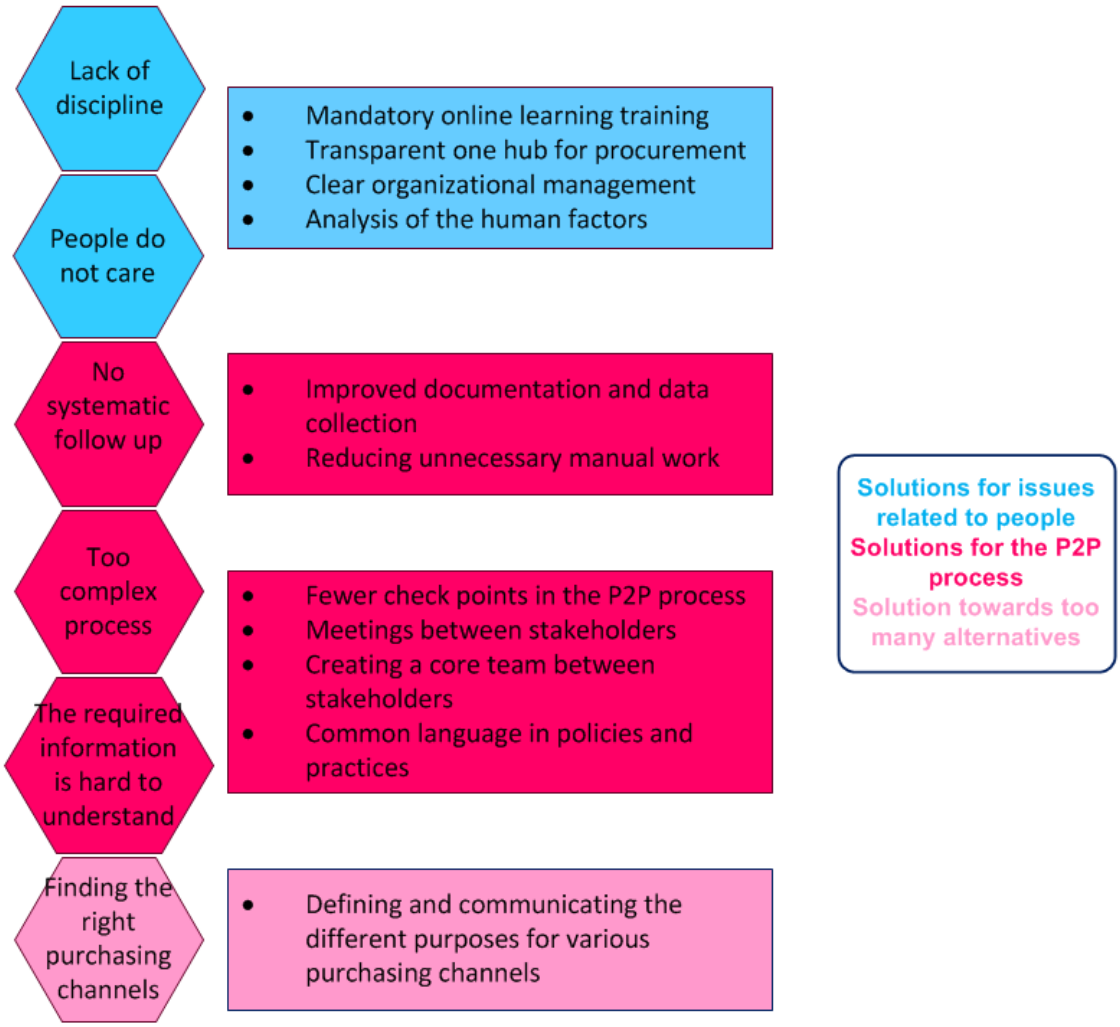



***Figure 26.*** *Summary of the suggestions for the case organization.*

The most beneficial improvement suggestion for the case organization would be to look deeper into the certain strategic vendors that create the poorest PO data quality. This means that it would be important to contact those employees who purchase from these vendors. The case organization could bring these issues for discussion to find out why the quality is poor also regarding the invoices. This would level up the data quality in an easy manner instantly. It would be important to understand, why the PO data is not sufficient and to provide further training. The case organization could look up also other countries, as the document analysis was only conducted in one country-level. This would bring more insight on the topic and it could be seen if there were certain vendors that have poor data quality in many countries or is it just country-specified. It would be also beneficial to look into maverick spend, if it has the same pattern or does it differ a lot. In addition, also a more vast and heterogeneous invoice and PO data should be investigated to have more insight on the topic. The case organization could analyze other data quality components, such as accuracy if the invoices and POs are mapped. In addition, looking into relevancy and timeliness might be useful.

## 7.5 Evaluation of the research

Saunders et al. (2000, p. 100-101) describes that the assessment should be conducted based on the study's credibility, dependability, confirmability and transferability. The credibility means how the reality reflects within the findings. Dependability on its own refers is the study coherent and consistent. The third factor, confirmability illustrates neutrality, subjectivity and objectivity in the research. Transferability of the study refers how well it can be applicable to another context. These dimensions and the assessment of these elements is picturized in figure 27.

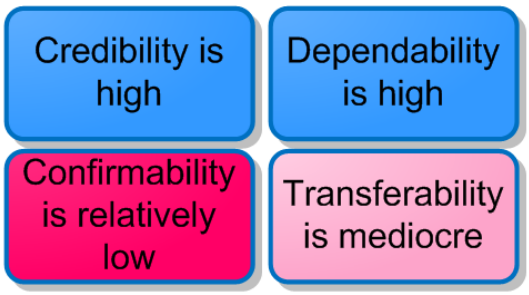



***Figure 27.*** *Evaluation of the research.*

The blue boxes represent that the study has been successful regarding credibility and dependability. On the other hand, transferability has been tolerable and confirmability only low.

### 7.5.1 Credibility and dependability of the study

Credibility is highly dependent on triangulation, which can be describes as using different information sources, researchers, theories and methods (Lincoln & Guba 1985). The wide range of these elements increases the validity of the research (Saunders et al. 2000, pp. 100-101). With quality triangulation, the probability of having a trustworthy content increases.

In this study, a variety of information sources were utilized. First, there were 17 theme interviews to build up the researcher's knowledge of the case organization and the situation. Although there were several people from different part of the organization to ensure that a wide range of opinions were heard, there is always a possibility that some perspectives were not taken into consideration. The people were from a wide range of positions in the organization: from experts to the executive team. The informality of these interviews increased the credibility, as it allowed the interviewees to mention issues out of a rigid scope. However, it has a downfall as the interpretation of these interviews in researcher-dependent. Nevertheless, commonly in a Master Thesis there is only one researcher and due to lack of resources this was a matter that was acknowledged during the

thesis process. Second, there were seven semi-structured interviews that were conducted precisely to get insight of chosen topic.

In addition, the researcher conducted a document analysis. The choice of vendors determined the whole analysis but the selection was decided based on the company's strategy. The invoices were straight from the vendors so it can be stated that they were reliable. Though some invoices had vendor's own product serial numbers or names, the researcher did not contact the vendor for more information. Even when the invoice amount was increased with vendors that had below 50 invoices in December 2016, the analysis did not differ much. This shows that the findings are consistent. However, it can be concluded that the differences might occur more if analyzed different vendors, rather than the same vendors with a larger time scale. The analysis of POs was more problematic as it was based on a report from ERP system. This stems from the fact, that the POs for vendors were not reliable because not every information is provided for the vendor due that it is internal data. When evaluating at the report, there could be errors or missing information if the data fields are not flowing correctly. At the same time, there were several reports to choose from so the selection of the report was a significant choice that steered the thesis' document analysis. This been said, in another report there could be information that would have been helpful for the study.

In the study, there was a thorough search of relevant academic literature. Mainly the researcher used books and scientific articles from prestigious journals to cover the research topic. The search was done across two different, separated areas to build a bridge in a niche-area. The purchasing process model challenged the theory part as due to its operational character there is not much literature regarding it. First, based on the theme interviews the theories were chosen from data and decision making perspective as well as procurement's process model. After this, there were the deeper interviews that defined and shaped the research scope. These two different themes were integrated to form an insightful theory part of a specified part of business. The decision making was added as a theory part and some other parts were removed. Moreover, there was a lack of literature considering spend management, spend data and spend data analysis. This might have had an impact on the thesis to be too narrow-minded. However, these arguments were discovered in the case organization too. It can be said that this thesis provided an updated view of spend management.

In addition, switching the topic for a more hands-on and approachable one, reflects that the study aims for a high credibility. It also has an impact on dependability as the study was documented relatively accurately to show all the steps of the process. As the research is a case study, the results can be seen as correct.

There were several methods utilized in this research to provide a robust cross-section. The empirical part had both quantitative and qualitative point of view so it had ambiguous

discovers. Alasuutari (2011, p. 231) mentions that a better understanding of the case organization is reached when there is both quantitative and qualitative research. As the qualitative research had a large sample, it provided diversity and it could be regarded more dependable as the patterns recurred. Nonetheless, there could have been a workshop among the procurement organization where the results would be evaluated to ensure that the results were understood in a correct manner.

### 7.5.2 Confirmability and transferability of the study

The process is highly affected of the interviews and of the other people from the organization that are involved. The objectivity was relatively well reached, as the researcher was only conducting the thesis in the case organization. In addition, the researcher did not know anyone from the organization beforehand. Moreover, the subject was also new so the study has been based merely on systematic research and not on presumptions. These elements reflect that the study has a relatively good confirmability regarding the situation.

The characters of case study disable widely to make generalizations in very high level because it focuses only one situation (Saunders et al. 2000, p. 101). This is a challenge in the eyes of transferability. The case organization's specific results cannot be generalized because they are centered merely in the case organization. It is also stated that for all researches it might not be a necessity to be applicable for multiple contexts (Alasuutari 2011, pp. 243-244). However, this research was conducted without industry-specific relations, so that it could be utilized in all kind of businesses. Furthermore, procurement is a part of the organization that is almost in every company so the function's key characters can be adapted to various businesses. For an example, the document analysis could be easily conducted in another organization with the same Pandit and Marmanis' (2001) defined lists of data attributes for POs and invoices.

## 7.6 Suggestions for further research

In this thesis, the spend data for price-quantity data was in higher quality in the invoice data set. It would be interesting to see does the relation between invoice and PO data stay the same, when looking at maverick spend; the spend that is small and distorted. After all, the spend data quality is lower in purchases that are not strategic priority and the control of the maverick spend might be sometimes not so rigidly managed. In addition, the impact of wild invoices should be taken into consideration, to see what kind of impacts the lack of no PO has.

As the subject is very operational and case-depended, a survey to interview different organizations would provide value. A best practice study would be beneficial to have hands-on information about different approaches to build spend analysis. Bringing the discus-

sion to the table could generate more researches and knowledge about the topic. The organizations could provide additional insights on how to manage or train people for better spend data, how the data sets are controlled and what kind of P2P model they use.

This study mainly covered the quality in the eye of completeness. It would be beneficial to examine other dimensions too to achieve a more insightful view on the quality as it consists of many dimensions. For instance, mapping invoices and POs could provide data on accuracy. Besides, timeliness and relevancy could be also looked from qualitative perspective to see how good quality the data is.

Third, the invoices and POs could be analyzed from a financial perspective, to achieve an integrated view on the topic due to the fact that procurement and finance are intertwined together. At the same time, a research about the financial process, for instance regarding invoicing and payments could be conducted. Another topic to study are the human factors that have a direct impact on data quality. Especially understanding the reasons for violating the process is key.

In the future, there could be a research based on merely vendors, not looking at the case organization. There could be comparisons between vendors – depending on the commodity, the number of items and money. In addition, the study could look into the purchasing complexity and how it affects the spend data quality.

The next step after correcting the spend data is to look how advanced technologies could provide value for spend data. This could be analyzed within few years when the technologies are more mature. For instance, the evaluation of the use of artificial intelligence that could learn to categorize the commodities right even if the data is wrong in the first place, would be a potential next step to be in the lead in procurement. Instead of manual processing, there could be rules that have to be fulfilled regarding the data fields, and if the requirements are met, the data could directly flow from the ERP system. There could be an analysis how much errors are reduced when there is less manual work, for example regarding POs. Aligning with this, neural networks can learn the human decision making process so the decisions made based on the spend data would be more error free.

# REFERENCES

Agmon, N. & Ahituv, N. 1987. Assessing data reliability in an information system. Journal of Management Information System. Vol. 8, Iss. 2.

Anttila, P. 1998. Tutkimisen taito ja tiedonhankinta. WWW-publications. Available (referenced on 28.6.2017): http://www.metodix.com/fi/sisallys/01_menetelmat/01_ tutkimuspro-sessi/02_tutkimisen_taito_ja_tiedon_hankinta/kooste#11.40.

Azvine, B., Cui, Z. & Nauck, D. D. 2005. Towards real-time business intelligence. BT Technology Journal. Vol. 23, Iss. 3.

Armando, G. 2012. Service productivity and service quality: A necessary trade off.

Ballou, D. P. & Pazer, H. L. Modeling data and process quality in multi-input, multi-output information systems. 1985. Management science. Vol. 31, Iss. 2, pp. 150.

Borthic, G. F. 2012. Designing Continuous Auditing for Highly Automated Procure-to-Pay Process. Journal of Information Systems. Vol. 26, Iss. 2, pp. 153-166.

Choo, C.W. 1996. The knowing organization: How organizations use information to construct meaning, create knowledge and make decisions. International journal of information management. Vol. 16, Iss. 5, pp. 329-340.

Choo, C. W. 2002. Information Management for the Intelligent Organization: The Art of Scanning the Environment. 3rd Edition. Medford, N.J.: Information Today Inc., p. 325.

Chowdary, P., Ettl, M., Dhurandhar, A., Ghosh, S. & Maniachari, G. 2011. Managing Procurement Spend using Advanced Compliance Analytics. 8th IEEE International Conference on e-Business Engineering.

Cox, A., Chicksand, D. & Ireland, P. 2005. Sourcing Indirect Spend: A Survey of Current Internal and External Strategies for Non-revenue-Generating Goods and Services. The Journal of Supply Chian Management. Vol. 40, Iss. 2, pp. 39-51.

Day, E. & Barksdale, H.C. Jr. 1994. Organizational purchasing of professional services: the process of selecting providers. The Journal of Business & Industrial Marketing, Vol. 9, Iss. 3, pp. 44-51.

Davenport, T. H. & Kirby, J. 2016. Just How Smart Are Smart Machines? MIT Sloan Management Review. Vol. 57, Iss. 3.

Davenport, T. H. & Snabe, J. H. 2011. How Fast and Flexible Do You Want Your Data, Really? MIT Sloan Management Review. Vol. 52, Iss. 3, pp. 57-62.

Eckerson, W. 2007. Best Practices in Operational BI. Converging Analytical and Operational Processes. TDWI Best Practices Report, Third Quarter 2007, 28 s.

Fitzsimmons, J. A., Noh, J., & Thies, E. 1998. Purchasing business services. Journal of Business and Industrial Marketing. Vol. 13, Iss. 4/5, pp. 370−380.

Hazen, B.T., Boone, C.A., Ezell, J.D. & Jones-Farmer, L.A. 2014. Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications. International Journal of Production Economics. Vol. 154, pp. 72-80.

Heikkilä, J., Vuori, M. & Laine, J. A. T. 2013. Purchasing Business Services. Teknologiainfo Teknova Oy. Helsinki, Finland.

Hirsjärvi, S., Remes, P., & Sajavaara, P. 2007. Tutki ja kirjoita (13th Edition). Helsinki, Kustannusosakeyhtiö Tammi. 448 p.

Hopkins, M. S. 2010. Are You Ready to Reengine Your Decision Making? MIT Sloan Management Review. Vol 52., nro 1 .

Hoyle, D. 2009. ISO 9000 Quality Systems Handbook - updated for the ISO 9001:2008 standard (Sixth Edition), Butterworth-Heinemann, Oxford, 2009, 802 p.

Gilad, T. & Gilad, B. 1986. SMR Forum: Business Intelligence – the Quiet Revolution. Sloan Management Review. Vol. 27, Iss. 4, pp. 53-61.

Grönroos, C. & Ojasalo, K. Service productivity towards a conceptualization of the transformation of inputs into economic results in services.

Käpylä, J. & Salonius, H. 2013. Tietojohtajan taskukirja: Tietojohtamisen näkökulmia aluekehittämiseen. Tampereen teknillinen yliopisto, Tietojohtamisen tutkimuskeskus NOVI, Tampere. 79 p.

Lee, R.G. & Dale, B.G. 1998. Business process management: a review and evaluation. Business Process Management Journal. Vol 4, Iss. 3, pp. 214-225.

Lee, Y.W., Strong, D. M., Kahn, B. K., Wang, R. Y. 2002. AIMQ: a methodology for information quality assessment. Information Management. pp. 133-146.

Loshin, D. 2013. Business Intelligence: The Savvy Manager's Guide. Morgan Kauffmann. Publishers, San Francisco, USA. 270 p.

Marzic, S., Krneta, P. & Pavlic, P. 2014. Spend analysis systems. 37th Internationl Convention on Information and Communication Technology, MIPRO.

Mukherjee, S., Fradkin, D., Roth, M. 2008. Classifying Spend Descriptions with off-the-shelf Learning components. Tools with Artificial integellince, 20th IEEE International Conference.

Olkkonen, T. 1994. Johdatus teollisuustalouden tutkimustyöhön. 2nd edition. Otaniemi, Espoo, 143 p.

Ojasalo, K., Moilanen, T., Ritalahti, J. 2014. Kehittämistyön menetelmät: Uudenlaista osaamista liiketoimintaan. 3rd edition. Sanoma Pro, Helsinki, 204 p.

Pandit, K. & Marmanis, H. 2008. Spend analysis: the window into strategic sourcing. J. Ross Publishing, P. 246

Pipino, L.L., Lee, Y.W. & Wang, R.Y. 2002. Data quality assessment. Communications of the ACM, Vol. 45, Iss. 4, pp. 211-218.

Porter, M.E. 1985. Competitive advantage: Creating and Sustaining Superior Performance. New York: Free Press, p. 557.

Porter, A. M. 1999. Taking control of 'indirect' corporate spending. Purchasing. Vol. 127, Iss. 3, pp. 55-60.

Redman, T. C. 1998. The impact of poor data quality on the typical enterprise. Communications of the ACM. Vol. 41, Iss. 2, pp. 79-82.

Rozemeijer, F., Quintens, L., Wetzels, M. & Gelderman, C. 2012. Vision 20/20: Preparing today for tomorrow's challenges. Journal of Purchasing & Supply Management. Vol 18, pp. 63-67.

Saunders, M., Lewis, P., & Thornhill, A. 2009. Research Methods for Business Students 5th Edition. England, Pearlson Education Limited. 614 p.

Still, A., Rhone, S. & Rosenbaum D. 2011. Driving Spend Management through Advanced Analytics. Supply Chain Management Review, Vol. 11, pp. 60-61.

Tayi, G.K. and Ballou, D.P. 1998. Examining data quality. Communications of the ACM, Vol. 41(2), pp. 54-57.

Thierauf, R. 2001. Effective Business Intelligence Systems. Greenwood Publishing-Groups, Westport, CT, USA, 370 s.

Thomas Jr., J. H. 2001. Business Intelligence – Why? eAI Journal, July, pp. 47-49.

Pirttimäki, V. & Hannula, M. 2003. Process models of business intelligence. Frontiers of e-business research.

Van der Valk, W. 2008. Service procurement in manufacturing: Results of three embedded case studies. Industrial Marketing Management. Vol. 37, Iss. 3, pp. 301-315.

Van der Valk, W. & Rozemeijer, F. 2009. Buying business services: towards a structured service purchasing process. Journal of Services Marketing. Vol. 23, Iss. 1, pp. 3-10.

Van der Valk, W. & Wynstra, F. Variety in business-to-business services and buyer-vendor interaction. International Journal of Operations & Production Management. Vol. 34, Iss. 2, pp. 195-220.

Vitt, E., Luckevich, M. & Misner, S. 2002. Business Intelligence: Making Better Decisions Faster. Washington: Microsoft Press, p. 202.

Wang, R.Y. & Strong, D.M. 1996. Beyond accuracy: What data quality means to data consumers. Journal of management information systems. Vol. 12, Iss. 4, pp. 5-33.

Wang, R.Y., Storey, V.C. & Firth, C.P. 1995. A framework for Analysis of Data Quality Research. IEE Transactions on Knowledge and data engineering, Vol. 7, Iss. 4, pp. 34.

Van Weele, A. 2005. Purchasing and Supply Chain Management.

Öykü, I., Jones, M. & Sidorova, A.. 2013. Business intelligence success: The roles of BI capabilities and decision environments. Information & Management, Vol. 50, pp. 13−23.